

VISIT...

LANZAROTE
Caliente.COM

KIRK-OTHMER

ENCYCLOPEDIA OF

CHEMICAL

TECHNOLOGY

EDITING - by **WATCHER**

VOLUME 7

FOURTH EDITION

COMPOSITE MATERIALS TO
DETERGENCY

CONTENT INDEX (vol 7)

(with hyperlinks)

Edited by <http://www.watcherworld.narod.ru/>

COMPOSITE MATERIALS	2
- SURVEY	3
POLYMER-MATRIX	17
- THERMOSETS	18
- THERMOPLASTICS	35
- CERAMIC-MATRIX	44
COMPUTER-AIDED DESIGN AND MANUFACTURING (CAD/CAM)	61
COMPUTER-AIDED ENGINEERING (CAE)	71
COMPUTER TECHNOLOGY	87
CONTACT LENSES	99
CONTRACEPTIVES	111
CONTROLLED RELEASE TECHNOLOGY	128
- AGRICULTURAL	129
- PHARMACEUTICAL	140
CONVEYING	153
COORDINATION COMPOUNDS	165
COPOLYMERS	176
COPPER	192
COPPER ALLOYS	216
- WROUGHT COPPER AND WROUGHT COPPER ALLOYS	217
- CAST COPPER ALLOYS	238
COPPER COMPOUNDS	253
COPYRIGHTS AND TRADEMARKS	262
- COPYRIGHTS	263
- TRADEMARKS	268
CORROSION AND CORROSION CONTROL	274
COSMETICS	285
COTTON	307
COUMARIN	319
CRYOGENICS	325
CRYSTALLIZATION	338
CUMENE	362
CYANAMIDES	366
CYANIDES	375
CYANINE DYES	389
CYANOCARBONS	403
CYANOHYDRINS	410
CYANURIC AND ISOCYANURIC ACIDS	417
CYCLOHEXANOL AND CYCLOHEXANONE	425
CYCLOPENTADIENE AND DICYCLOPENTADIENE	429
DAIRY SUBSTITUTES	438
DATABASES	453
DEFOAMERS	462
DENTAL MATERIALS	470
DENTIFRICES	501
DESICCANTS	505
DESIGN OF EXPERIMENTS	519
DETERGENCY	526

COMPOSITE MATERIALS

Survey,
Polymer-matrix,
Ceramic-matrix,

SURVEY

The term *composite material* is used to describe macroscopic combinations of two or more materials. Macroscopic combinations are specified to exclude alloys that consist of materials combined on a microscopic scale (1). Such an exclusive definition of composite materials is not universal, but it is commonly accepted and it helps restrict to a manageable size an introductory treatment of the science and technology of composite materials.

The fundamental goal in the production and application of composite materials is to achieve a performance from the composite that is not available from the separate constituents or from other materials. The concept of improved performance is broad and includes increased strength or reinforcement of one material by the addition of another material. This is the well-known purpose in the alloying of metals and in the incorporation of chopped straw into clay for bricks by the ancient Egyptians and plant fibers into pottery by the Incas and Mayans. These ancient productions of composite materials consisted of reinforcing brittle materials with fibrous substances. In both cases the mechanics of the reinforcement was such as to reduce and control the production of cracks in the brittle material during fabrication or drying (2).

The modern interest in composite materials can be traced to the development of Bakelite, or phenolic resin, in 1906. Bakelite was a hard, brittle material that had few if any mechanical applications on its own. However, the addition of a filler—the earliest applications used short cellulose fibers (2)—yielded Bakelite molding compounds that were strong and tough and found early applications in mass-produced automobile components. The wood flour additive improved Bakelite's processibility and physical, chemical, and electrical properties, as well as reducing its cost (3,4).

In 1920, experiments showed that bulk glass, which has a tensile strength of about 200 MPa (29,000 psi), could be produced in fiber form with strengths of about 4 GPa (580,000 psi). The strengths of the glass fibers increased rapidly as the glass was drawn to smaller diameters. It was also observed that the strength of the glass fibers decreased with time, implying that the glass fibers were very sensitive to surface defects. The incorporation of strong glass fibers in plastic matrices produced practical bulk materials in which the plastic matrix functioned as a binder, transferring load to the fibers and separating the fibers. The fibers were thus protected and prevented from coming in contact with one another and possibly reducing their strength by abrasion. The term *reinforced plastics*, commonly used for these materials, is, therefore, misapplied (2,6), inasmuch as it is the plastic that maintains the strength of the fiber, rather than the fiber that reinforces the plastic. Early applications of glass-reinforced plastics date back to the end of World War II, when radomes were fabricated by wet lay-up of yarns of glass woven into cloth and impregnated with polyester resin. Glass-reinforced plastics are still used for radomes in aerospace and submarine applications. Large volumes of chopped strand mat (CSM) consisting of short lengths of glass fibers (25 to 75 mm) randomly arranged have been used in marine and automotive applications.

In the early 1960s there was a significant increase in interest in the science of composite materials when very strong and stiff, but brittle, ceramic, boron, and carbon fibers became available. These new fibers were at least as strong as the glass fibers of the same diameter but approximately five to ten times stiffer. The term *advanced composites* has been coined to cover those reinforced plastics containing continuous strong, stiff fibers such as carbon (graphite), boron, aramid, or glass. Advanced composites have been developed primarily for the aerospace industry, in which the demand for strong and stiff lightweight structures overcame the prohibitive costs of early composite material systems. Currently, advanced composite materials, mainly carbon-reinforced epoxy resins, are used widely in military aircraft and increasingly in civil aircraft. The largest volume usage of these materials has, however, been in the recreational area in applications such as tennis rackets and golf club shafts (7).

Composite Classification and Terminology

Composites are generally classified as fibrous, laminated, and particulate (1).

Fibrous Composites. These composites consist of fibers in a matrix. The fibers may be short or discontinuous and randomly arranged; continuous filaments arranged parallel to each other; in the form of woven rovings (collections of bundles of continuous filaments); or braided (8). In the case of chopped strand mat the random arrangement is planar. In whisker (needle-shaped crystals or filaments of carbon and ceramics) reinforced materials the arrangement is usually three-dimensional and the resulting composites are macroscopically homogeneous.

Laminates. Two or more layers of material bonded together form a laminated composite. Common examples of laminates are in automobile windshields (laminated glass) and bimetal thermostats (9). In both cases homogeneous, isotropic layers of materials are bonded together to form nonhomogeneous composite laminates (see LAMINATES).

Plastic laminated sheets produced in 1913 led to the formation of the Formica Products Company and the commercial introduction, in 1931, of decorative laminates consisting of a urea–formaldehyde surface on an unrefined (kraft) paper core impregnated with phenolic resin and compressed and heated between polished steel platens (8,10). The decorative surface laminates are usually about 1.6 mm thick and bonded to wood (a natural composite), plywood (another laminate), or particle board (a particulate composite). Since 1937, the surface layer of most decorative laminates has been fabricated with melamine–formaldehyde, which can be prepared with mineral fillers, thus offering improved heat and moisture resistance and allowing a wide range of decorative effects (10,11).

Plywood is a laminate consisting of thin sheets of wood arranged with the grain in alternate sheets at right angles (12) (see BUILDING MATERIALS, SURVEY; WOOD-BASED COMPOSITES AND LAMINATES). Invented in the mid-nineteenth century using natural based adhesives, the manufacture of plywood was greatly improved by the introduction of phenolic resins (10,11). Indeed, plywood accounts for more than 60% of the total phenolic resin production (13).

Fiber composite laminates often consist of unidirectional (parallel) continuous fibers in a polymer matrix with the individual layers, plies or laminae, stacked with selected fiber angles so as to produce specific laminate stiffness and strength values. Laminates are also made by stacking layers of mats or woven fabric, which have many possible weave configurations (14,15). The arrangement of the layers of a laminate is known as the stacking sequence. Laminates can be classified as unidirectional, in which case all, or nearly all (8), of the fibers lie in the same direction (7). A schematic of a unidirectional laminate is shown in Figure 1. Unidirectional laminates are orthogonal with respect to their principal axes which are parallel to the fibers (1) and perpendicular to the fibers (2), as indicated in Figure 1. Unidirectional laminae can be stacked so as to tailor the laminate to produce a wide range of mechanical responses. A commonly used code for specifying the stacking sequence of a laminate uses the subscript T for the total stacking sequence, starting from the top surface, and the subscript S for a laminate that is symmetric about the midplane (7). The laminate shown in Figure 2 could be referred to as $[0/\theta/0]_T$ or $[0/\theta \bar{I}]_S$, where the bar above the θ denotes a ply that is symmetrically placed about the middle surface of the laminate. A common stacking sequence is cross-ply, in which the fibers in each ply are at 0° or 90° ; for example, the laminate $[0/90]_S$ is a four-ply symmetric cross-ply laminate with the top and bottom surface fiber directions at 0° (aligned to some reference direction), and the middle two plies at 90° . Another common laminate uses the angle-ply stacking sequence, in which the fiber orientations are $\pm\theta$. Often the plies are balanced so that there are equal numbers of $+\theta$ and $-\theta$ plies. The laminate $[45/-45]_S$ has a balanced symmetric angle-ply stacking sequence. If the reference direction were rotated by 45° , then this angle-ply laminate would be represented as the cross-ply laminate with the stacking sequence $[0/90]_S$. The laminate $[45/-45/0/90]_T$ is also balanced, but not symmetric (7). Both the cross-ply and angle-ply laminates are examples of bidirectional laminates (7). Quasi-isotropic laminates are multidirectional laminates that have two equivalent (isotropic) elastic constants in the plane of the laminate (1,6,7). Examples of quasi-isotropic laminates are $[0/60/-60]_T$, $[0/60/-60]_S$, and $[0/45/-45/90]_S$. Even though the first example is not a symmetric laminate, it is defined as quasi-isotropic because the definition refers

only to the in-plane stiffness properties (1). In all the laminates mentioned so far the thickness of each ply is the same. Such laminates are sometimes referred to as regular laminates. Special laminate properties can be obtained by using asymmetric (16) or unbalanced laminates, as in the case of the Grumman X-29A aircraft, which has a forward-swept wing configuration in which the laminate is tailored to counteract (through coupling between flexure and torsion) the aeroelastically unstable effects of the upward twisting tendency of the wing (17).

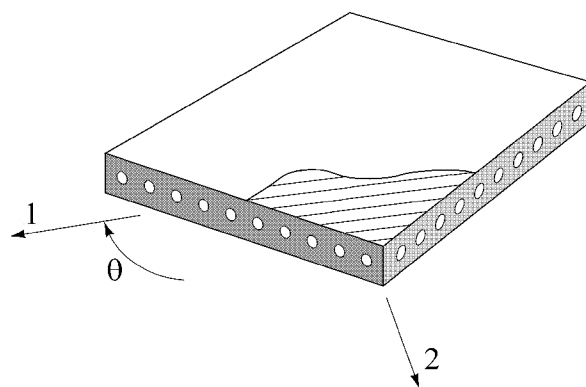


Fig. 1. A unidirectional lamina. The material axes are labeled 1 for the fiber direction and 2 for the direction transverse to the fibers.

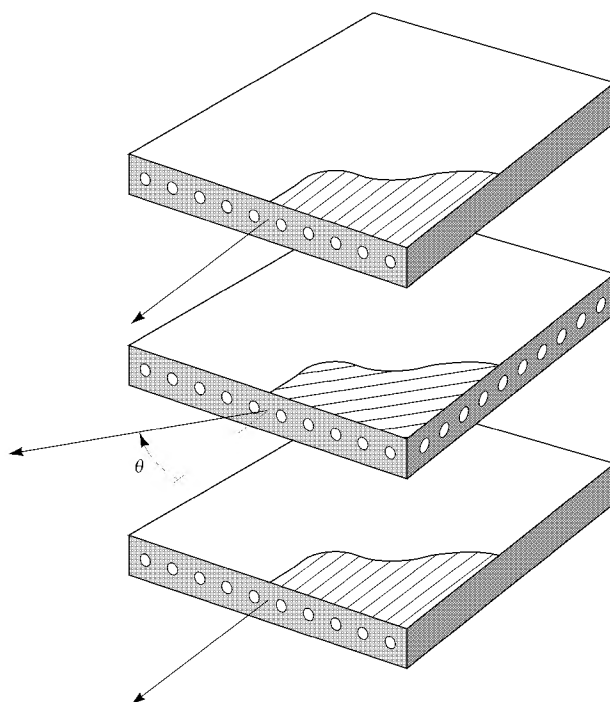


Fig. 2. Three layers of unidirectional material stacked together to form a laminate.

Particulate Composites. These composites encompass a wide range of materials. As the word *particulate* suggests, the reinforcing phase is often spherical or at least has dimensions of similar order in all directions. Examples are concrete, filled polymers (18), solid rocket propellants, and metal and ceramic particles in metal matrices (1).

Concrete consists of an aggregate of coarse (>5 mm) or fine particles (gravel, crushed rock, sand, slag, etc) bound in a cement matrix, with the aggregate content constituting typically 66–78% by weight (19). In concrete, the role of the aggregate has often been regarded as that of an inert filler, introduced to reduce cost (20). However, significant technical improvements in volume stability, durability (20), and strength and stiffness (18,20) are obtained with concrete in comparison with cement (qv).

In polymeric materials particles are used as fillers to improve strength, toughness, processibility, dimensional stability, frictional wear and lubrication properties, and, in some cases, resistance to ultraviolet radiation (21,22). The particle characteristics can be classified as sphere, cube, block, flake, and fiber (21). Common fillers (qv) are glass (spheres), calcite (cube and block), kaolin, mica and talc (flakes), and wood flour (flake) (21). The scale of the particles used depends on the role and ranges from submicrometer to several hundred micrometers. In some cases, the filler can improve some properties while degrading others, as in nylon-6,6, which has improved strength and stiffness but reduced impact strength when filled with 30% glass (23).

Solid rocket propellants represent a very special case of a particulate composite in which inorganic propellant particles, about 75% by volume, are bound in an organic matrix such as polyurethane. An essential requirement is that the composite be uniform to promote a steady burning reaction (1). Further examples of particulate composites are those with metal matrices and include cermets, which consist of ceramic particles in a metal matrix, and dispersion hardened alloys, in which the particles may be metal oxides or intermetallic compounds with smaller diameters and lower volume fractions than those in cermets (1). The general nature of particulate reinforcement is such that the resulting composite material is macroscopically isotropic.

Reinforcements

The choice of reinforcement for a particular engineering application is likely to depend on a large number of parameters, including strength, stiffness, environmental stability, long-term characteristics, and cost. At present, carbon, glass, and aramid fibers account for over 95% of the industrial market, with increasing use in areas such as the aerospace, automotive, construction, biomedical, and sport sectors. No single fiber type can be said to be truly superior to another; each has its own merits as well as shortcomings. The stress–strain responses obtained by bending typical carbon, glass, and aramid fibers are summarized in Figure 3. Carbon fibers generally offer the highest strengths and stiffnesses but can be brittle, failing at relatively low applied strains. Glass fibers offer intermediate strengths and higher failure strains but exhibit lower moduli. Aramid fibers such as Kevlar, a proprietary material developed by Du Pont, have lower strengths but are capable of absorbing considerable energy without fracture. Rather than using strength and stiffness as the performance parameters for material selection, specific properties obtained by normalizing the property with respect to the density (ρ) of the fiber are often quoted. For

tensile or compressive members, the specific strength (σ/ρ) yields the largest load-carrying capacity for a given mass and the specific stiffness (E/ρ) produces the smallest deflection for a given mass (18). A convenient way of comparing fibers is to plot specific strength against specific modulus, as shown in Figure 4. The data presented in Figure 4 consist of the properties of the fiber types discussed in the following sections, with the values for steel wire also shown for comparison. Because some of the fibers are extremely sensitive to flaws, the values of the specific strengths may be subject to wide variations.

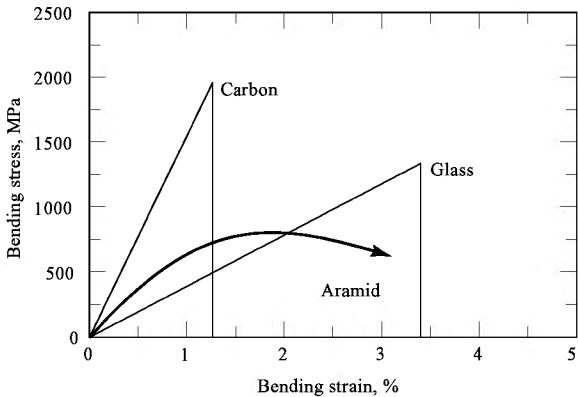


Fig. 3. Bending stress versus bending strain for typical carbon, glass, and aramid fibers. To convert MPa to psi, multiply by 145.

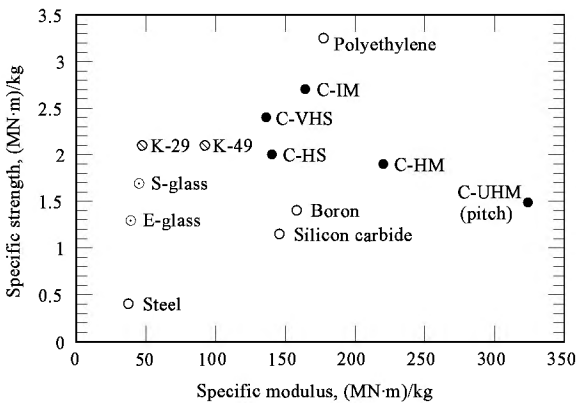


Fig. 4. Specific properties of reinforcing fibers for composite materials. K-29 and K-49 represent Kevlar aramid fibers. Other designations are explained more fully in the text. To convert (MN·m)/kg to (lbf·in.)/lb, multiply by 4.03×10^6 .

The following sections also include brief presentations of manufacturing procedures because the final properties of the fibers depend on the processing conditions. For more about the structure or mechanical characteristics of the fibers, the reader is directed to more comprehensive discussions presented elsewhere in the *Encyclopedia* or in the literature (8,24–30).

Glass Fibers. Glass fibers represent the most frequently used reinforcement in modern polymer composites. This popularity results from their relatively low cost and high tensile strength. In bulk form, glass is brittle, having a relatively low strength. However, when extruded and drawn into fine fibers, the strength of glass increases enormously, by as much as two orders of magnitude. Glass-fiber-reinforced composites are currently used in a wide variety of applications, including boat construction and the automotive and aerospace industries. Many types of glass fiber are available, each having specific properties and characteristics. The most commonly used fibers, E-glass, contain approximately 14% Al_2O_3 , 18% CaO , 5% MgO , 8% B_2O_3 , and 1% $\text{Na}_2\text{O} + \text{K}_2\text{O}$ as well as silica (see GLASS).

Glass fibers are produced by dry mixing the individual components and then heating them to form a melt. The temperature of the melt depends on the glass composition but is typically about 1250°C. The molten glass then passes into the fiber-drawing furnace and subsequently flows through a large number of tiny orifices forming fine filaments. By pulling these filaments swiftly and continuously, the glass can be drawn into fine circular fibers with a diameter typically between 10 and 13 μm . In this form, glass fibers can be damaged very easily; a slight touch can cause a significant reduction in strength. It is common practice, therefore, to apply a size, a surface coating, to protect the fibers during handling and prevent damage during any subsequent processing stages, such as weaving. The size needs to be carefully formulated so as to hold the individual fibers together and afford sufficient lubrication when they come in contact with the processing equipment. Once the fibers have been woven into their final form, a coupling agent is applied. The coupling agent reacts with the fiber and matrix phases of the composite material to promote a strong bond at the interface.

Standard E-glass fibers are inexpensive to manufacture and offer satisfactory mechanical properties, such as a modulus of 72 GPa (1.04×10^7 psi) and a tensile strength of about 3.45 GPa (500,000 psi) (27). E-glass fibers are used extensively in automotive and marine applications, in the sports industry, and in chemical plants. Certain applications, however, require higher modulus fibers, which can be achieved by incorporating greater quantities of Al_2O_3 and MgO into the basic silica compound. Higher modulus, 87 GPa (1.26×10^7 psi), and strength, 4.6 GPa (667,000 psi), are achieved with S-glass fibers (27), which are used in the aerospace sector, where the increased production cost is offset by the significant reduction in weight of parts manufactured using such fibers. Other types of glass fibers that offer superior resistance in acidic and alkaline environments, radiation protection, and also dielectric characteristics are also available, albeit at a higher cost than the basic E-glass fiber.

Other advantages of glass fibers include their good energy-absorbing characteristics when used in composite form and their ability to resist very high temperatures; also, they do not burn, do not absorb water, have a low coefficient of thermal expansion, and offer excellent electrical insulation. Their greatest limitation is, however, their relatively low modulus. One means of overcoming this limitation is to combine glass fibers with a higher modulus fiber, for example, carbon fibers, so that the merits of both fibers combine to produce a strong, stiff, hybrid composite material.

Carbon Fibers. Although fine carbon filaments were first manufactured from a cellulose base material (8) as long ago as 1880, modern production and manufacturing of carbon fibers began in significant quantities in the early 1960s. The mechanical properties offered by carbon fiber composites, such as high strength and stiffness, are directly attributable to the highly anisotropic nature of the carbon crystal (see CARBON AND GRAPHITE FIBERS). The current technology for manufacturing carbon fibers is based on the thermal decomposition of organic precursor materials such as polyacrylonitrile (PAN) and pitch. Early efforts aimed at developing processes using rayon precursors were found to produce a low yield of carbon fibers with relatively low strength and modulus (28). Generally, the stages in the production of carbon fibers from the various organic precursors can be identified as spinning, stabilization, carbonization, and graphitization. Most carbon fibers are based on the PAN precursor, which is first pumped through a spinneret into a coagulating bath. The spinneret contains between 1000 and 320,000 holes producing tows of continuous fibers (24). Stretching the tow after spinning has the effect of increasing the strength and modulus of the carbon fibers (30). These intermediate modulus (IM) fibers have finer diameters than

the unstretched high strength (HS) fibers, as a result of the postdrawing process, and moduli and strengths of up to 300 GPa (4.35×10^7 psi) and 5.2 GPa (754,000 psi), respectively. In the stabilization process, needed to convert the precursor into a thermally stable form capable of withstanding the high temperatures and processing rates with a good yield (28), the PAN fibers are oxidized in air at 200–300°C, under strain. The carbonization stage involves rapid pyrolysis in an inert atmosphere, typically nitrogen, at temperatures between 1000 and 1500°C. This procedure yields the previously mentioned high strength fibers, which have a modulus of about 230 GPa (3.3×10^7 psi) and a strength of about 3.3 GPa (478,500 psi). The strength of the PAN-based carbon fibers reaches a maximum at a carbonization temperature of about 1500°C (24,29,31). In this stage of processing almost all elements other than carbon are eliminated as volatiles, with some residual amounts of nitrogen, perhaps, and stretched ring carbon structures are produced (28). High modulus (HM) fibers can be produced by subsequently heating in a nitrogen or argon atmosphere to a temperature between 2000 and 3200°C in the graphitization stage. For carbon fibers produced using rayon precursors, the graphitization is usually performed with the fibers under strain, making the process expensive and, as a result of the relatively low yield, inefficient.

Although the terms *graphite* and *carbon* are often used interchangeably, they are also used to distinguish graphite fibers, which have been subjected to heat treatment at temperatures greater than 1700°C, have some degree of preferred orientation, and have tensile moduli of the order of 345 GPa, from carbon fibers, which have lower degrees of preferred orientation, heat-treatment temperatures, and tensile moduli (8). The term *graphite fibers* can also refer to carbon fibers that are heated at temperatures above 2000°C (30).

Fibers produced from pitch precursors can be manufactured by heat treating isotropic pitch at 400 to 450°C in an inert environment to transform it into a liquid crystalline state. The pitch is then spun into fibers and allowed to thermoset at 300°C for short periods of time. The fibers are subsequently carbonized and graphitized at temperatures similar to those used in the manufacture of PAN-based fibers. The isotropic pitch precursor has not proved attractive to industry. However, a process based on anisotropic mesophase pitch (30), in which commercial pitch is spun and polymerized to form the mesophase, which is then melt spun, stabilized in air at about 300°C, carbonized at 1300°C, and graphitized at 3000°C, produces ultrahigh modulus (UHM) carbon fibers. In this process tension is not required in the stabilization and graphitization stages.

Surface treatment of fibers is an important stage in the manufacturing process. The primary aim of such treatment is to improve the adhesion between the fiber and the matrix and to improve handleability (29,32). The types of treatment are oxidative gas or dry oxidation, wet oxidation, and aqueous electrolytic or anodic oxidation. Anodic oxidation, which uses electrolytes such as sulfuric acid, potassium sulfate, or sodium hydroxide and oxidation times of 1 to 10 min, is the main industrial process (32). Oxidation treatments involving temperatures of about 2800°C (8) have led to significant increases in fiber strength by removing volume and surface flaws, thus resulting in the production of very high strength fibers (VHS). Finally, an organic coating, or size, is applied to the fibers. Often this coating is the same as the matrix used in the composite (29).

Pitch-based fibers generally have higher moduli but lower strengths than their PAN-based counterparts. The specific properties of the various types of carbon fibers are compared in Figure 4. Pitch-based fibers also have higher electrical conductivity, which can be an important consideration in certain circumstances, for example, for use in electromagnetic inductance (EMI) shielding.

Carbon-fiber-reinforced plastic composites are currently used in primary and secondary aircraft structures, helicopter rotor blades, sporting goods such as fishing rods and tennis rackets, and certain biomedical applications. As well as offering superior stiffness and strength, carbon fibers offer good thermal stability, a low and negative (-0.5 to -1.2×10^{-6} /°C) axial coefficient of thermal expansion (6), and excellent fatigue resistance when used in composite form. Laminated composites based on carbon fibers suffer certain limitations, perhaps the most significant being their relatively low energy-absorbing capacity and, therefore, their poor resistance to transverse impact loading. They also suffer poor abrasion resistance, are attacked by certain acids, and undergo galvanic corrosion when brought into contact with certain metals and alloys.

Aramid Fibers. Aromatic polyamide fibers exhibiting a range of mechanical properties are available from several manufacturers, perhaps the best known being Du Pont's proprietary fiber Kevlar. These fibers possess many unique properties, such as high specific tensile strength and modulus (see Fig. 4). Aramid fibers have good chemical resistance to water, hydrocarbons, and solvents. They also show excellent flame retardant characteristics (see HIGH PERFORMANCE FIBERS; POLYAMIDES).

Aramid fibers are formed by mixing a polymer, poly(*p*-phenylene terephthalamide), with a strong acidic solution and extruding the mixture through spinnerets at a temperature between 50 and 100°C. The fibers are cooled and then washed thoroughly before being dried on bobbins. The properties of the aramid fibers can be modified by using solvent additives or by using a postspinning heat treatment. Aramid fibers with relatively low moduli, such as Kevlar 29 (K-29 in Fig. 4), find use in energy-absorbing applications such as bullet-resistant and other protective clothing, helmets, and ropes. The higher modulus counterpart, Kevlar 49 (K-49 in Fig. 4), is used in high performance engineering applications such as the manufacture of load-bearing components for the aerospace industry where it is used in filament-wound rocket motor cases (33).

Kevlar fibers are highly crystalline in structure, with rigid molecular chains oriented in their longitudinal direction (27). The weak hydrogen bonds in the transverse direction result in a highly anisotropic fiber with low transverse tensile strengths. Consequently, when heavily loaded, such fibers tend to split extensively or to fibrillate. This is clearly a limitation in certain applications. Other limitations of these fibers include their low compressive strength, sometimes poor adhesion to polymer matrices, and tendency to absorb water. Special tooling is needed for cutting or machining.

Other Fibers. Boron fibers attracted considerable attention in the 1960s and 1970s as a result of their high modulus and strength (approximately 400 GPa and 3.5 GPa, respectively). Each filament of boron fibers consists of a core of either tungsten wire or carbon fiber on which a layer of boron is applied (34). The boron is vapor deposited onto the central filament from a boron trichloride–hydrogen solution. The resulting fibers have a large cross section, typically between 100 and 200 μm , and, therefore, offer very high stability in compressive stress fields. Boron fiber–epoxy composites were used in the stabilizers of the Grumman F-14 and McDonnell Douglas F-15 aircraft (1,24). Currently boron fibers are more commonly used in metal matrices such as aluminum and magnesium. Boron-reinforced plastics, however, still find use as repair patches for damaged aircraft structures as well as in certain sporting goods, such as fishing rods and golf club shafts (24).

Polyethylene fibers are attracting considerable interest owing to their high specific strength and stiffness as well as their excellent energy-absorbing capability. Two methods are used to manufacture polyethylene fibers (see FIBERS, OLEFIN). The first involves the extrusion and drawing of a medium molecular weight polyethylene. The second method involves dissolving the polymer in a solvent and spinning at a temperature of around 140°C. The fibers so produced offer tensile strengths up to 3.2 GPa (464,000 psi) and moduli as high as 170 GPa (2.5×10^7 psi) (24). The specific strength and modulus of high performance polyethylene fibers compare very favorably with those of aramid and carbon fibers (see Fig. 4). However, their wider application is limited by relatively poor compressive properties and a low melting point, 150°C. The possibility of hybridizing polyethylene fibers with stiffer carbon fibers is under investigation.

Silicon carbide fibers exhibit high temperature stability and, therefore, find use as reinforcements in certain metal matrix composites (24). Silicon fibers have also been considered for use with high temperature polymeric matrices, such as phenolic resins, capable of operating at temperatures up to 300°C. Silicon carbide fibers can be made in a number of ways, for example, by vapor deposition on carbon fibers. The fibers manufactured in this way have large diameters (up to 150 μm), and relatively high Young's modulus and tensile strength, typically as much as 430 GPa (6.2×10^7 psi) and 3.5 GPa (507,500 psi), respectively (24,34) (see REFRACTORY FIBERS).

Matrix Materials

The mechanical properties of composites based on the fibers discussed depend not only on the characteristics of the fibers but also on those of the matrix itself as well as on the fiber–matrix interface.

Polymer Matrices. The matrix in a polymer composite serves both to maintain the position and orientation of the fibers and to protect them from potentially degrading environments. Polymer matrices may be thermosets or thermoplastics. Thermosetting polymers are rigid, cross-linked materials that degrade rather than melt at high temperatures; thermoplastics are linear or branched molecules that soften upon heating. A comparison of the mechanical properties and relative cost of various thermosetting and thermoplastic materials is given in Table 1 (35), and the critical composite property, interlaminar fracture toughness G_{Ic} , is shown in Figure 5 (36,37). Thermoset-based composites are somewhat less expensive than thermoplastic-based composites, but have lower heat distortion temperatures and poorer toughness when tested in an interlaminar mode (36). The majority of present-day

composite components are still based largely on thermosetting matrices, such as unsaturated polyesters, epoxies, and phenolic resins (see COMPOSITE MATERIALS, POLYMER-MATRIX).

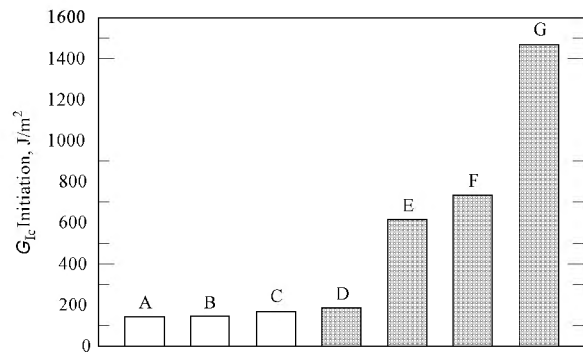


Fig. 5. Interlaminar fracture toughness, G_{IC} , for a number of thermosetting and thermoplastic composites (36,37). Open white bars represent glass-fiber composites; shaded bars are for carbon fibers. The materials are A, polyester (unidirectional); B, vinyl ester (CSM = chopped strand mat); C, epoxy (R/BR1424); D, epoxy (T300/914); E, PPS; F, PES; and G, PEEK. To convert J/m^2 to $ftlb/in.^2$ multiply by 2100.

Table 1. Mechanical Properties and Relative Costs of Thermosetting and Thermoplastic Composites

Matrix material	Young's modulus GPa ^a	Tensile strength, MPa ^b	Heat distortion temp, °C	Relative cost ^c
Thermosets				
polyester	3.6	60	95	1
vinyl ester	3.4	83	110	1.8
epoxy	3.0	85	110	2.3
phenolic	3.0	50	120	0.8
Thermoplastics				
PES	2.8	84 ^d	203	6
PEI	3.0	105 ^d	200	7
PEEK	3.7	92 ^d	140	25

^a To convert GPa to psi, multiply by 145,000.

^b To convert MPa to psi, multiply by 145.

^c Relative to the 1988 cost of isophthalic polyester (35).

^d Yield stress.

Thermosetting Matrices. Unsaturated polyester resins represent one of the most commonly used matrix systems employed in the manufacture of composite components. The marine industry in particular has profited greatly from the moderate cost and ease of use of polyester resins. Several types of polyester resins are presently available, offering a wide range of properties such as superior strength or resistance to acidic environments. The curing reaction of polyester resins is exothermic, providing heat to further accelerate the curing process. Composites based on polyester resins are generally manufactured using hand lay-up or spray-up techniques. Large, single-piece components, such as boat hulls and radomes, can be manufactured using polyesters, although care has to be taken when manufacturing thick laminates to ensure that the exothermic heat generation does not damage or degrade the composite. Standard polyester resins offer lower tensile strengths than many of the other thermosets and are limited in terms of their upper operating temperature (see POLYESTERS, UNSATURATED).

Vinyl ester resins generally offer mechanical properties superior to those of polyester matrices but at an increased cost. Vinyl esters are chemically similar to epoxy resins but are manufactured via a cold-curing process similar to that used in the manufacture of polyester resins. Vinyl esters offer superior resistance to water and chemical attack and are used in such applications as underground pipes, tank liners, and storage tanks (see VINYL POLYMERS).

Phenolic resins (qv), once a popular matrix material for composite materials, have in recent years been superseded by polyesters and epoxies. Nevertheless, phenolic resins still find considerable use in applications where high temperature stability and fire resistance are of paramount importance. Typical examples of the use of phenolic resins in the marine industry include internal bulkheads, decks, and certain finishings. The curing process involves significant production of water, often resulting in the formation of voids within the volume of the material. Further, the fact that phenolics are prone to absorb water in humid or aqueous conditions somewhat limits their widespread application. Phenolic resins are also used as the adhesive in plywood, and phenolic molding compounds have wide use in household appliances and in the automotive, aerospace, and electrical industries (12).

As a result of their increased cost, epoxy resins (qv) tend to be used as matrices for high performance continuous fiber composites. Epoxies offer mechanical properties and water resistance superior to those of polyesters, as well as lower shrinkage during cure. The cure process involves the addition of a hardener and possibly an accelerator as well as a temperature cycle between 60 and 180°C. The resulting mechanical characteristics of the resin depend on the cross-link density of the polymer. High temperature matrices, such as those used in the aerospace sector, tend to have higher cross-link densities and are therefore more brittle. The toughness of epoxy resins can be increased by reducing the cross-link density or by adding a lower modulus phase, such as carboxyl terminated butadiene acrylonitrile (CTBN) rubber or thermoplastic additives. Epoxy-based composites have been used widely in aircraft structures such as rudders, torsion boxes, spoilers, engine pylon fairings, and in racing car bodies. Although very stiff composites result when epoxy resins are combined with carbon fibers, their resistance to localized impact loading is often poor. Indeed, relatively low energy impacts, such as those associated with a dropped tool, can introduce large areas of delamination and thereby significantly reduce residual strength. Repair technology for thermoset-based composites is not very advanced, and a greater understanding of the significance of various defects on load-bearing ability is required.

For higher temperature applications, thermosetting addition polyimides (qv) are becoming increasingly important. Polyimide matrix materials were developed in an attempt to produce composites capable of withstanding temperatures of 316°C (600°F) (38,39). One such system is the PMR-15 (*in situ* polymerization of monomer reactants) polyimide developed at NASA Lewis Research Center. Cure temperatures tend to be higher and cure cycles tend to be longer so that the cost of polyimide systems is greater than that of epoxies. Composites made from polyimide prepreg are generally brittle at room temperature, although significant improvements in toughness are being achieved in new material formulations. For aerospace applications polyimides have attractive hot-wet properties, but special attention has to be paid to the fiber-matrix interface and the fiber sizing employed (24). Applications of

PMR-15–carbon fiber composites include jet-engine cowlings, ducts, compressor blades, and flaps and fairings (24,38).

Thermoplastic Matrices. Many of the thermoplastic-based composites such as carbon-fiber-reinforced polyetheretherketone (PEEK) offer excellent resistance to impact loading (40) and are thereby suitable for use in high performance engineering applications. Other interesting aspects of these composites include the possibility of thermoforming and shaping at elevated temperature and the potential for thermal joining and repair, as well as recycling. Long fiber thermoplastic composites can be manufactured directly from prepregged sheets of fibers by a film-stacking sequence in which layers of fibers and film are stacked alternately, using plies of commingled polymer filaments and reinforcing fibers, or they can be manufactured from thermoplastic rovings such as thermoplastic impregnated fibers (FIT).

Carbon fiber–PEEK is a semicrystalline thermoplastic matrix composite. The base polymer is extremely expensive compared with many of the thermosetting matrices, so the PEEK composites are being considered for use in high performance engineering parts such as aerospace components (see POLYETHERS). PEEK is a very ductile aromatic polymer (40) with a strain to failure in excess of 100%. Composites based on this polymer are capable, therefore, of absorbing considerable energy before incurring damage. This quality is most evident under impact conditions where delaminated areas for a given incident energy may be three times smaller than in comparable epoxy-based composites (41,42). Carbon fiber–PEEK also offers high interlaminar fracture toughness (see Fig. 5), as well as superior fatigue properties. Large areas of delamination such as those induced by impact loading can be repaired successfully within short periods of time (43). One of the disadvantages of PEEK-based composites, however, is that they require high processing temperatures, typically 380 to 400°C. Also, it is suggested that a rapid cooling rate be employed in order to ensure that the resulting degree of crystallinity is not too high (44).

Poly(phenylene sulfide) (PPS) is another semicrystalline polymer used in the composites industry. PPS-based composites are generally processed at 330°C and subsequently cooled rapidly in order to avoid excessive crystallization and reduced toughness. The superior fire-retardant characteristics of PPS-based composites result in applications where fire resistance is an important design consideration. Laminated composites based on this material have shown poor resistance to transverse impact as a result of the poor adhesion of the fibers to the semicrystalline matrix. A PPS material more recently developed by Phillips Petroleum, AVTEL, has improved fiber–matrix interfacial properties, and promises, therefore, an enhanced resistance to transverse impact (see POLYMERS CONTAINING SULFUR).

A number of amorphous thermoplastics are presently employed as matrices in long fiber composites, including polyethersulfone (PES), polysulfone (PSU), and polyetherimide (PEI). All offer superior resistance to impact loading and higher interlaminar fracture toughnesses than do most epoxies. However, the amorphous nature of such polymers results in a lower solvent resistance, clearly a limitation if composites based on such polymers are to be used in aggressive environments.

Other Matrix Materials. Advanced materials, eg, structural components, in aerospace vehicles also employ ceramics and metals as composite matrices (see COMPOSITE MATERIALS, CERAMIC-MATRIX; METAL-MATRIX COMPOSITES).

Fabrication Methods

A large number of methods are presently available for manufacturing long fiber composites. The cost of finished components depends not only on the price of the raw materials but also on labor costs and energy requirements. Many composite components are manufactured by hand, involving relatively long manufacturing cycles. Large engineering components, such as boat hulls, are frequently manufactured using hand lay-up or spray-up techniques. The whole cycle may last several months, involving a large work force. Considerable effort is being made to find cheaper, more efficient ways to manufacture composite parts and structures. Some of the more commonly employed techniques used to manufacture such parts are outlined in the following. For more information, the reader is directed to References 24 and 25.

Hand and Spray Lay-Up. Hand lay-up techniques are used quite extensively in the marine industry for constructing relatively small numbers of large components. Typically, parts are manufactured using a female mold onto which a gel coat is applied. The structure can then be built up using sheets of glass fiber plies or mats and applying a polymer resin (typically a catalyzed polyester) with a brush. For large structures, the fibers and resin are dispensed from a gantry that spans the width of the structure. Composites with fiber weight fractions between 25 and 45% can be manufactured using this procedure.

A similar technique, spray-up molding, involves spraying short fibers (25 to 50 mm in length) mixed with a catalyzed resin onto the mold. The glass–resin mixture is then consolidated by manually rolling the surface. Typically, spray lay-up results in laminates with fiber weight fractions of between 25 and 30%. Advantages of the hand and spray lay-up techniques include low equipment and tooling costs, the ability to include inserts, and the ability to manufacture sandwich constructions. Both procedures have drawbacks, including high labor costs, low production rates, and poor reproducibility, as well as the fact that only one surface has a high quality finish.

Filament Winding. Bodies of revolution such as cylinders and spherical shells can be manufactured economically by filament winding (24). Continuous fiber bundles or tows are wound onto a rotating mandrel. The fibers may be impregnated with resin (most frequently a thermoset) just before being wound onto the mandrel (wet winding) or may be prepregged with a partially cured matrix (dry winding). The positioning of the fibers is achieved by a feeder arm located in front of the rotating mandrel. The wrap angle, that is the angle between the fiber and the axis of the mandrel, can usually be varied between 0 and almost 90°, thereby enabling multidirectional components to be realized relatively easily. Once the winding process is finished the whole fixture is cured in an oven in order to yield the desired mechanical properties and structural integrity. The mandrel is then removed by dissolving (for salt- and sand-based mandrels), melting (for low melting point alloys), or collapsing and dismantling (for steel mandrels). Recently, attempts have been made to filament wind thermoplastic-based composites. The prepregged tow or roving is heated by a laser or a hot-air blower just before being wound onto the mandrel. Although this procedure is still in its infancy and more work is required in order to optimize the technique, it does offer great potential because the need for a subsequent curing process is removed. Filament winding requires a considerably greater financial investment than do conventional hand lay-up techniques. However, increased expense is in part offset by the greater reproducibility of the resulting parts as well as by the higher production rate.

Parts with fiber volume fractions up to 60% can be fabricated by filament winding. The procedure is often used to manufacture composite rocket motors, corrosion-resistant tanks and storage containers, and piping for below-ground applications.

Autoclave Molding. Bag molding methods include vacuum bag, pressure bag, and autoclave molding (45). Autoclave molding is well suited to the manufacture of large, high quality components. The autoclave itself is a cylindrical pressure vessel capable of generating pressures up to several atmospheres. Heating of the autoclave and, thereby, the composite is achieved by a series of electrical heaters. The composite part is placed against the mold surface and backed with porous PTFE film and a bleeder layer made of glass-fiber cloth, as shown in Figure 6. This layer serves to absorb any excess resin that escapes from the composite during curing. The whole construction is then covered by a vacuum membrane enabling a vacuum to be created within the stack. The cure cycle for a typical epoxy-based composite (Fiberdux 913) is (1) the autoclave is heated to 90°C at between 2 and 5°C/min, and a vacuum of 75 kPa (560 mm Hg) is applied; (2) the temperature is maintained at 90°C for 30 min in order to allow resin flow; (3) 700 kPa (6.9 atm) pressure is applied to the autoclave; (4) the vacuum is vented when the pressure reaches 140 kPa (1.4 atm); (5) the temperature is increased to 120°C and maintained for 1 h. During this stage gelation and curing take place. (6) The part is cooled and is removed when it falls below 90°C.

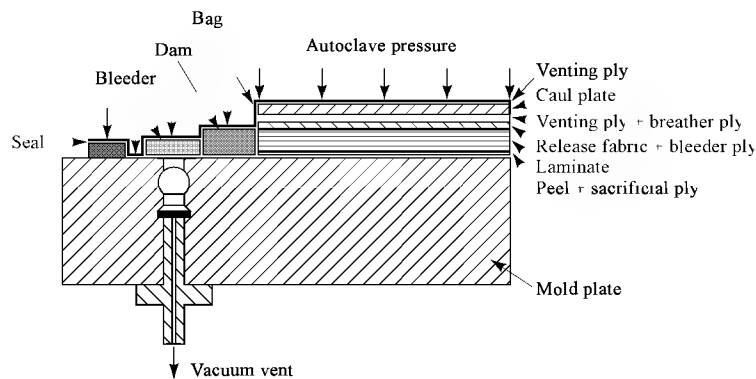


Fig. 6. Autoclave molding of a thermoset-based composite.

The processing cycle is long and relatively expensive. Autoclave molding is used most commonly, therefore, for manufacturing limited high quality components, such as parts for the aerospace industry. The cross section of an autoclave-molded carbon fiber composite is shown in the micrograph in Figure 7 in which the interface between plies of different fiber orientation are clearly visible and are associated with resin-rich regions.

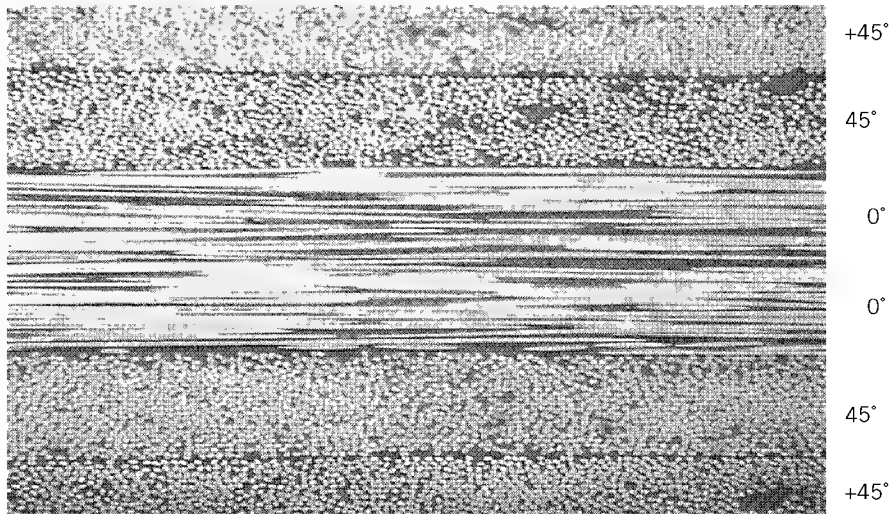


Fig. 7. Micrographic cross section of an autoclaved carbon fiber composite.

Compression Molding. This molding method is used to manufacture small-and intermediate-sized high quality parts at relatively high rates. A typical flash mold is shown schematically in Figure 8. Flash is the term used to refer to the extruded material that is typically found along the parting lines of a mold (8). The desired number of layers of preimpregnated composite are placed in the lower mold, and the upper mold is lowered until the required pressure is achieved. The composite is then heated according to the required cycle. Certain polymers, such as epoxy resins, flow during processing. For this reason, flash grooves are often machined into the perimeter of the lower mold to catch any excess resin. In order to achieve a high quality surface, great care has to be taken in the manufacture of the mold surfaces. After continued use, the surfaces need to be replaced in order to maintain the desired tolerances. In spite of the costs involved in the production and maintenance of the mold, compression molding represents a relatively inexpensive and practical means of producing high quality composite parts.

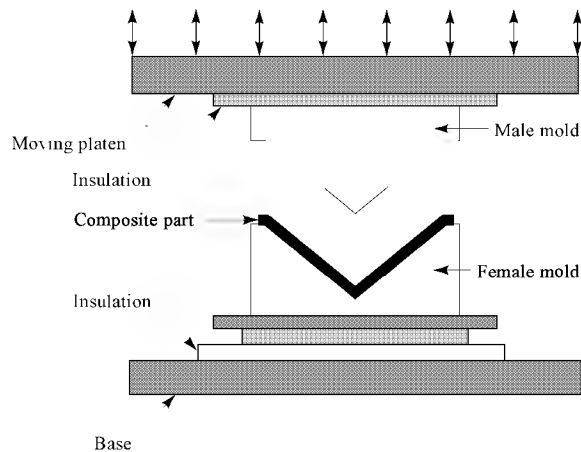


Fig. 8. Compression molding in matched dies.

Pultrusion. This technique is commonly used for producing continuous sections of unidirectionally reinforced polymers (46) (Fig. 9). Typically, continuous fibers are pulled through a resin bath and then through a series of wiper rings that serve to remove any excess resin. The fiber–resin mixture is then pulled through a spider, which separates and evenly distributes the fibers before passing into the die. Once in the die, the material is heated, resulting in gelation and curing of the resin. On leaving the die, the composite is sufficiently hard and strong to be pulled mechanically to a large-diameter collecting drum or to a cutter wheel. All the commonly used continuous fibers, such as carbon, glass, and Kevlar, can be pultruded. Pultrusion of hybrid composites is also possible. Thermosetting resins such as polyester and epoxy are commonly used; thermoplastic matrix composites can also be pultruded (47).

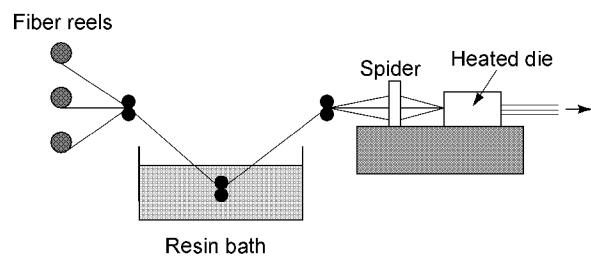


Fig. 9. Schematic representation of pultrusion process for making continuous fiber rods.

Resin Transfer Molding. This molding technique (RTM) allows the economical manufacture of high quality composites. Dry fibers, which may have the form of continuous strand mat, unidirectional, woven or knitted preforms, are placed in a closed mold and resin is then introduced into the mold under external pressure or vacuum. The resin cures under the action of its own exotherm, or heat may be applied to the mold to complete the curing process. Early applications of the resin transfer molding technique employed unsaturated polyester resins. Polyester and vinyl ester resins are used in resin-transfer-molded consumer products, pipes, pressure vessels, and automotive applications. Epoxy resins have also been developed for resin transfer molding of high quality, high fiber volume fraction composite components in electronic and aerospace applications (48).

Theories of Reinforcement

The advantages of composites include improved stiffness and strength of the composite material system compared with the base-line or unreinforced material. For example, the introduction of straw into bricks and plant fibers into pottery by early civilizations resulted in improved strength of the bricks and pottery by preventing cracking induced in drying out of the brittle (crack-sensitive) materials. Brittle materials can be strengthened by increasing toughness. In composites, enhanced toughness can be achieved by increasing the energy required to initiate and propagate a crack through the brittle matrix. Increased work of fracture has been obtained in composites through debonding at the fiber–matrix interface; frictional interaction between the fiber and the matrix as fibers bridging a matrix crack are pulled out of the matrix as the crack extends and opens; deformation of the fibers; and fiber fracture. In the case of brittle matrix composites, small quantities of reinforcing material are needed to achieve the desired performance.

Fiber volume fraction is a quantitative measure of degree of reinforcement of the matrix material in a fiber-reinforced composite. If the volume of a composite material is V and the volume of the fibers is V_f and that of the matrix is V_m then

$$V = V_f + V_m$$

Similarly for weights W , W_f and W_m

$$W = W_f + W_m$$

The density ρ of the composite material is then related to the density of the fiber ρ_f and matrix ρ_m by the rule of mixtures:

$$\rho = v_f \rho_f + v_m \rho_m$$

where

$$v_f = \frac{F_f}{V}$$

and

$$v_m = \frac{V_m}{V}$$

are the fiber and matrix volume fractions, respectively. Thus, for carbon fibers with a density of 1.8 g cm^{-3} in an epoxy resin with a density of 1.2 g cm^{-3} , the density of the composite as a function of the fiber volume fraction is given in Figure 10.

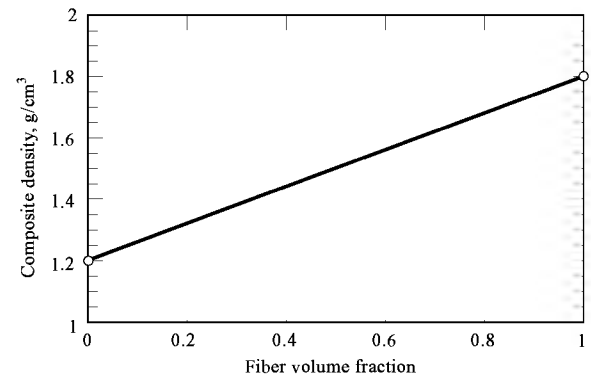


Fig. 10. The variation of the density of carbon-fiber reinforced epoxy resin with the fiber volume fraction, based on the rule of mixtures.

The stiffness of E_1 of a unidirectional lamina in the fiber direction is also given approximately by the rule of mixtures:

$$E_1 = v_f E_f + v_m E_m$$

For the case of high modulus fibers such as carbon fibers with $E_f = 240 \text{ GPa}$ ($3.5 \times 10^7 \text{ psi}$), in a polymer matrix, such as epoxy resin with $E_m = 3.0 \text{ GPa}$ (450,000 psi), the extensional modulus is approximately proportional to the fiber volume fraction and the modulus of the fibers:

$$E_1 \approx v_f E_f$$

Thus the addition of the stiff carbon fibers has a very great effect in stiffening the epoxy matrix. For the commonly used fiber volume fraction of 0.6 the unidirectional carbon–epoxy lamina has a predicted extensional stiffness $E_1 = 145 \text{ GPa}$ ($2.1 \times 10^7 \text{ psi}$).

The relationship between fiber and matrix moduli and fiber volume fraction for a unidirectional lamina loaded in the direction transverse to the fibers is not simple. A lower bound (1) is given by the expression of the series spring model,

$$\frac{1}{E_2} = \frac{v_f}{E_f} + \frac{v_m}{E_m}$$

and an upper bound is provided by the rule of mixtures, or parallel spring model. Generally, the fibers may be anisotropic, so that only the transverse fiber modulus should be used. Since this property is difficult to measure (6), it is usual to employ the axial fiber properties and assume isotropic behavior of the constituents in the micromechanics models. A simple model has been developed for the extensional modulus of fibrous and particulate composites using a combination of the series and parallel spring models (49):

$$\frac{E_2}{E_m} = \frac{X}{1 + X\left(1 - \frac{E_m}{E_f}\right)} + (1 - X)$$

where X depends on the reinforcement and packing geometry. For example, for fibers of circular cross section in a square array,

$$\sqrt{\frac{2v_f}{\pi}} < X < \sqrt{\frac{4v_f}{\pi}}$$

An estimate of the shear modulus G_{12} is also given by an expression based on the series spring model

$$\frac{1}{G_{12}} = \frac{v_f}{G_f} + \frac{v_m}{G_m}$$

The variation of the in-plane shear modulus normalized with respect to the matrix modulus as a function of the fiber volume fraction is shown in Figure 11. As noted earlier, it is generally difficult to measure the shear modulus of the fibers, which may themselves be anisotropic. The equation should be used with caution.

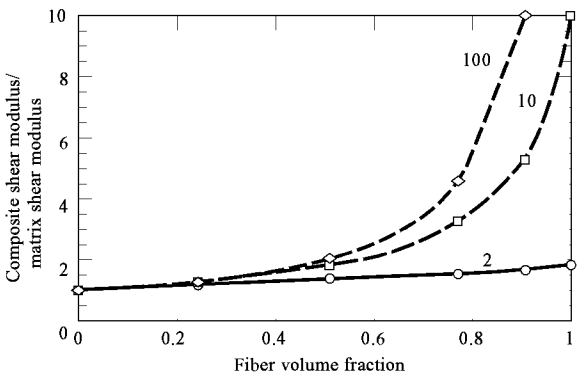


Fig. 11. The variation of the shear modulus G_{12} of carbon-fiber-reinforced epoxy resin as a function of the fiber volume fraction for several values of the ratio of the fiber shear modulus to that of the matrix (G_f/G_m). Ratios are noted on the curves (100,10,2).

Poisson's ratio, ν_{12} , is the negative of the ratio of the strain transverse to the fiber direction, ϵ_2 , and the strain in the fiber direction, ϵ_1 , when the lamina is loaded in the fiber direction and can also be expressed in terms of the properties of the constituents through the rule of mixtures.

$$\nu_{12} = v_f \nu_f + v_m \nu_m$$

The difference between the bounds defined by the simple models can be large, so that more advanced theories are needed to predict the transverse modulus of unidirectional composites from the constituent properties and fiber volume fractions (1). The Halpin-Tsai equations (50) provide one example of these advanced theories in which the rule of mixtures expressions for the extensional modulus and Poisson's ratio are complemented by the equation

$$\frac{M}{M_m} = \frac{1 + \xi \eta v_f}{1 - \eta v_f}$$

where

$$\eta = \frac{\left(\frac{M_f}{M_m}\right) - 1}{\left(\frac{M_f}{M_m}\right) + \xi}$$

M is the composite property (E_1, G_{12}, ν_{21}), and M_f and M_m are the corresponding fiber and matrix properties. The principal problem in the application of the Halpin-Tsai equations lies in determining the value of ξ , which depends on the fiber and packing geometry and loading conditions (1).

Strength predictions of composites are in general quite complex and somewhat limited. This is particularly true of compressive and shear strengths, which are needed, together with the tensile strengths, in composite failure prediction.

The tensile strength of a unidirectional lamina loaded in the fiber direction can be estimated from the properties of the fiber and matrix for a special set of circumstances. If all of the fibers have the same tensile strength σ_f and the composite is linear elastic until failure of the fibers, then the strength of the composite is given by

$$\sigma_c = \sigma_f v_f + \epsilon_{mf} E_m v_m$$

where ϵ_{mf} is the strain in the matrix when the fibers fail. For carbon fibers in epoxy resin the tensile strength of the composite is predicted to be approximately proportional to the fiber volume fraction. The assumption of a constant failure stress of the fibers is unrealistic. This strength prediction relates to a fiber-dominated model. However, at low values of fiber volume fraction the fiber-dominated model is invalid. This is seen clearly in the limiting case of zero fiber volume fraction when the matrix strength σ_m is observed. For very low values of fiber volume fraction the composite tensile strength is given approximately by the matrix-dominated model:

$$\sigma_c = \sigma_m (1 - v_f)$$

Both tensile strength equations are illustrated in Figure 12 using data typical of graphite fibers in epoxy resin.

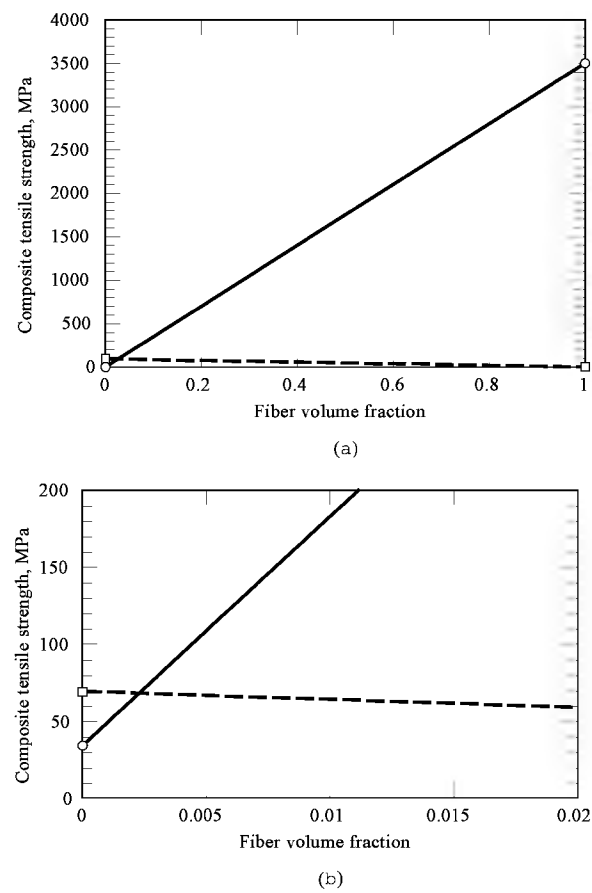


Fig. 12. (a) The variation of the tensile strength of unidirectional carbon-fiber-reinforced epoxy resin as a function of the fiber volume fraction. (b) The variation of the tensile strength of unidirectional carbon-fiber-reinforced epoxy resin as a function of the fiber volume fraction for low fiber volume fractions.

The four quantities E_1 , E_2 , G_{12} , and ν_{12} are sufficient to define the stress–strain law for a unidirectional or woven fiber ply under plane stress, loaded in the principal material directions.

$$\begin{Bmatrix} \sigma_1 \\ \sigma_2 \\ \tau_{12} \end{Bmatrix} = \begin{bmatrix} \frac{E_1}{1 - \nu_{12}\nu_{21}} & \frac{\nu_{12}E_2}{1 - \nu_{12}\nu_{21}} & 0 \\ \frac{\nu_{12}E_2}{1 - \nu_{12}\nu_{21}} & \frac{E_2}{1 - \nu_{12}\nu_{21}} & 0 \\ 0 & 0 & G_{12} \end{bmatrix} \begin{Bmatrix} \epsilon_1 \\ \epsilon_2 \\ \gamma_{12} \end{Bmatrix}$$

where

$$\nu_{21} = \frac{\nu_{12}E_2}{E_1}$$

This stress–strain relationship is often written as (1)

$$\begin{Bmatrix} \sigma_x \\ \sigma_y \\ \tau_{xy} \end{Bmatrix} = \begin{bmatrix} Q_{11} & Q_{12} & 0 \\ Q_{12} & Q_{22} & 0 \\ 0 & 0 & Q_{66} \end{bmatrix} \begin{Bmatrix} \epsilon_x \\ \epsilon_y \\ \gamma_{xy} \end{Bmatrix} \quad \text{or} \quad (\sigma) = [Q](\epsilon)$$

where $[Q]$ is the reduced stiffness matrix.

The stress–strain relationship is used in conjunction with the rules for determining the stress and strain components with respect to some angle θ relative to the fiber direction to obtain the stress–strain relationship for a lamina loaded under plane strain conditions where the fibers are at an angle θ to the loading axis. When the material axes and loading axes are not coincident, then coupling between shear and extension occurs and

$$\begin{Bmatrix} \sigma_x \\ \sigma_y \\ \tau_{xy} \end{Bmatrix} = \begin{bmatrix} Q_{11} & Q_{12} & Q_{16} \\ Q_{12} & Q_{22} & Q_{26} \\ Q_{16} & Q_{26} & Q_{66} \end{bmatrix} \begin{Bmatrix} \epsilon_x \\ \epsilon_y \\ \gamma_{xy} \end{Bmatrix}$$

or

$$(\sigma) = [\bar{Q}](\epsilon)$$

in which the transformed stiffness terms are related to the reduced stiffnesses and angle θ by (1)

$$\begin{aligned} Q_{11} &= Q_{11} \cos^4 \theta + 2 (Q_{12} + 2 Q_{66}) \sin^2 \theta \cos^2 \theta + Q_{22} \sin^4 \theta \\ Q_{12} &= (Q_{11} + Q_{22} - 4 Q_{66}) \sin^2 \theta \cos^2 \theta + Q_{12} (\sin^4 \theta + \cos^4 \theta) \\ Q_{22} &= Q_{11} \sin^4 \theta + 2 (Q_{12} + 2 Q_{66}) \sin^2 \theta \cos^2 \theta + Q_{22} \cos^4 \theta \end{aligned}$$

$$\begin{aligned} Q_{16} &= (Q_{11} \quad Q_{12} \quad 2 \, Q_{66}) \sin\theta \cos^3\theta + (Q_{12} \quad Q_{22} + 2 \, Q_{66}) \sin^3\theta \cos\theta \\ Q_{26} &= (Q_{11} \quad Q_{12} \quad 2 \, Q_{66}) \sin^3\theta \cos\theta + (Q_{12} \quad Q_{22} + 2 \, Q_{66}) \sin\theta \cos^3\theta \\ Q_{66} &= (Q_{11} + Q_{22} - 2 \, Q_{12} - Q_{66}) \sin^2\theta \cos^2\theta + Q_{66} (\sin^4\theta + \cos^4\theta) \end{aligned}$$

Classical laminated plate theory is used to determine the stiffness of laminated composites. Details of the Kirchhoff-Love hypothesis on which the theory is based can be found in standard texts (1,7,51). Essentially, the strains in each ply of the laminate are represented as middle surface strains ($\epsilon_x^\circ, \lambda_y^\circ, \gamma_{xy}^\circ$) plus a component that is proportional to the curvatures and twist ($\kappa_x, \kappa_y, \kappa_{xy}$) and the distance, z_k the ply is from the middle surface (Fig. 13).

$$\begin{Bmatrix} \epsilon_x \\ \epsilon_y \\ \gamma_{xy} \end{Bmatrix} = \begin{Bmatrix} \epsilon_x^\circ \\ \epsilon_y^\circ \\ \gamma_{xy}^\circ \end{Bmatrix} + z \begin{Bmatrix} \kappa_x \\ \kappa_y \\ \kappa_{xy} \end{Bmatrix}$$

The resultant forces (N_x, N_y, N_{xy}) and moments (M_x, M_y, M_{xy}) are determined by integrations of the stresses in each ply.

$$\begin{Bmatrix} N_x \\ N_y \\ N_{xy} \end{Bmatrix} = \int_{-t/2}^{t/2} \begin{Bmatrix} \sigma_x \\ \sigma_y \\ \tau_{xy} \end{Bmatrix}_k \, dz$$

and

$$\begin{Bmatrix} M_x \\ M_y \\ M_{xy} \end{Bmatrix} = \int_{-t/2}^{t/2} \begin{Bmatrix} \sigma_x \\ \sigma_y \\ \tau_{xy} \end{Bmatrix}_k \, z \, dz$$

where k denotes the k th ply located at a distance z_k from the laminate middle surface and t is the thickness of the laminate (Fig. 13). The stiffness of the laminate is then given as

$$\begin{Bmatrix} N_x \\ N_y \\ N_{xy} \end{Bmatrix} = \begin{bmatrix} A_{11} & A_{12} & A_{16} \\ A_{12} & A_{22} & A_{26} \\ A_{16} & A_{26} & A_{66} \end{bmatrix} \begin{Bmatrix} \epsilon_x^\circ \\ \epsilon_y^\circ \\ \gamma_{xy}^\circ \end{Bmatrix} + \begin{bmatrix} B_{11} & B_{12} & B_{16} \\ B_{12} & B_{22} & B_{26} \\ B_{16} & B_{26} & B_{66} \end{bmatrix} \begin{Bmatrix} \kappa_x \\ \kappa_y \\ \kappa_{xy} \end{Bmatrix}$$

and

$$\begin{Bmatrix} M_x \\ M_y \\ M_{xy} \end{Bmatrix} = \begin{bmatrix} B_{11} & B_{12} & B_{16} \\ B_{12} & B_{22} & B_{26} \\ B_{16} & B_{26} & B_{66} \end{bmatrix} \begin{Bmatrix} \epsilon_x^\circ \\ \epsilon_y^\circ \\ \gamma_{xy}^\circ \end{Bmatrix} + \begin{bmatrix} D_{11} & D_{12} & D_{16} \\ D_{12} & D_{22} & D_{26} \\ D_{16} & D_{26} & D_{66} \end{bmatrix} \begin{Bmatrix} \kappa_x \\ \kappa_y \\ \kappa_{xy} \end{Bmatrix}$$

where the coefficients of the laminate stiffness matrices are given in terms of the ply stiffnesses as

$$(A_{ij}, B_{ij}, D_{ij}) = \int Q_{ij}(1, z, z^2) \, dz$$

or, for a laminate with N plies,

$$\begin{aligned} A_{ij} &= \sum_{k=1}^N (Q_{ij})_k (z_k - z_{k-1}) \\ B_{ij} &= \frac{1}{2} \sum_{k=1}^N (Q_{ij})_k (z_k^2 - z_{k-1}^2) \\ D_{ij} &= \frac{1}{3} \sum_{k=1}^N (Q_{ij})_k (z_k^3 - z_{k-1}^3) \end{aligned}$$

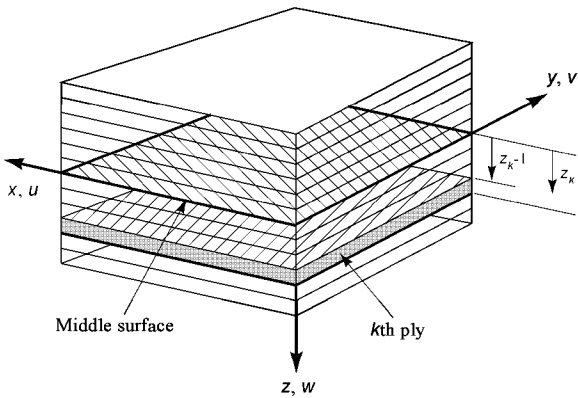


Fig. 13. Schematic of a laminate with the coordinates and ply notation used in laminated plate theory.

An important observation is that generally there will be coupling between extension and bending. A laminate that is symmetric about the middle surface does not suffer from this coupling because the terms of the $[B]$ matrix are zero. A further coupling problem, that of shear-extension coupling, can arise in the $[A]$ matrix, as it did in the lamina stiffness matrix $[Q_i]$, if the terms A_1 and A_2 are not zero. This coupling is eliminated in balanced laminates that have equal numbers of plies of orientations θ and $-\theta$. Thus the majority of laminates are fabricated as balanced symmetric laminates. Coupling between bending and twisting can also be problematic. It is not generally possible to achieve zero values of D_1 and D_2 , so it is usual to ensure that these terms are small by using laminates with large numbers of alternating layers (6,51).

Quasi-isotropic laminates have the same in-plane stiffness properties in all directions (1), which are defined in terms of the $[A]$ matrix of the laminate. For the laminate to be quasi-isotropic,

$$A_{11} = A_{22} \quad \text{and} \quad A_{11} - A_{12} = 2 A_{66}$$

and the laminate has two equivalent elastic constants (7). For the case of a quasi-isotropic laminate with highly isotropic plies the equivalent elastic constants can be expressed as

$$E \approx \frac{3}{8}E_1 + \frac{5}{8}E_2 \quad \text{and} \quad G \approx \frac{1}{8}E_1 + \frac{1}{4}E_2$$

These approximations can also be used for laminates in which the fibers are randomly oriented, provided that the modulus of the fibers is much greater than that of the matrix.

The strength of laminates is usually predicted from a combination of laminated plate theory and a failure criterion for the individual lamina. A general treatment of composite failure criteria is beyond the scope of the present discussion. Broadly, however, composite failure criteria are of two types: noninteractive, such as maximum stress or maximum strain, in which the lamina is taken to fail when a critical value of stress or strain is reached parallel or transverse to the fibers in tension, compression, or shear; or interactive, such as the Tsai-Hill or Tsai-Wu (1,7) type, in which failure is taken to be when some combination of stresses occurs. Generally, the ply materials do not have the same strengths in tension and compression, so that five-ply strengths must be determined:

- X the tensile strength parallel to the fibers
- X' the compressive strength parallel to the fibers
- Y the tensile strength in the plane of the ply perpendicular to the fibers
- Y' the compressive strength in the plane of the ply perpendicular to the fibers
- S the in-plane shear strength

These values are determined by experiment. It is, however, by no means a trivial task to measure the lamina compressive and shear strengths (52,53). Also the failure of the first ply of a laminate does not necessarily coincide with the maximum load that the laminate can sustain. In many practical composite laminates first-ply failure may be accompanied by a very small reduction in the laminate stiffness. Local ply-level failures can reduce the stress-raising effects of notches and enhance fatigue performance (54).

Economic Considerations

In the form of fiber-reinforced unidirectional and multidirectional composites very high values of strength and stiffness can be achieved with fiber volume fractions of about 60%. Excellent fatigue properties can also be obtained (54). Transverse impact damage tolerance, which was an early limiting factor in carbon fiber composites, has been improved greatly through the development of high strain to failure fibers and tough matrix and interleaf materials. However, these improvements in properties have been associated with significant increases in the material costs. There have been many applications in which cost has been a secondary factor to performance, such as in the military aerospace fields (24). However, cost has become the critical issue in the continued development and application of advanced composite materials. Composite components can be fabricated through various routes depending on the quality of the end product, and economies in the finished cost can be made by reducing the number of parts and attachments and assembly operations. In comparing costs it is important to include the life-cycle costs. In the oil and gas industry significant reductions in the life-cycle costs can be achieved by replacing steel with fiber-reinforced plastics. Another very large market with enormous potential for the application of medium technology composites is the automotive industry.

BIBLIOGRAPHY

"Composite Materials," in *ECT* 3rd ed., Vol. 6, pp. 683–700, by W. B. Hillig, General Electric Co.; "Composites, High Performance," in *ECT* 3rd ed., Suppl. Vol., pp. 260–281, by J. Davis, NASA.

1. R. M. Jones, *Mechanics of Composite Materials*, Scripta Book Co., Washington, D.C., 1975.
2. J. E. Gordon, *The New Science of Materials*, Penguin Books Limited, Middlesex, UK, 1971.
3. R. A. Baker and co-workers in S. M. Lee, ed., *International Encyclopedia of Composites*, Vol. 2, V. C. H. Publishing, Inc., 1990, pp. 182–222.
4. J. A. Brydson, *Plastics Materials*, 5th ed., Butterworths, London, 1989, Chapt. 7.
5. A. A. Griffith, *Philos. Trans. R. Soc. London. Ser. A* **A221**, 163–198 (1921).
6. F. P. Gerstle, Jr., in J. I. Kroschwitz, ed., *High Performance Polymers and Composites*, John Wiley & Sons, Inc., New York, 1991, pp. 83–127.
7. S. W. Tsai, *Composites Design*, 4th ed., Think Composites, Dayton, Ohio, 1988.
8. G. Lubin, ed., *Handbook of Composites*, Van Nostrand Reinhold Co., Princeton, N.J., 1982.
9. U. U. Savolainen and R. M. Sears, in A. G. H. Dietz, ed., *Composite Engineering Laminates*, MIT Press, Cambridge, Mass., 1969, Chapt. 10, pp. 212–226.
10. A. DiNoto, *Art Plastics*, Cross River Press, New York, 1984.
11. S. Katz, *Classic Plastics*, Thames and Hudson, London, 1984.
12. J. F. Keegan, in I. Rubin, ed., *Handbook of Plastic Materials and Technology*, John Wiley & Sons, Inc., New York, 1990, Chapt. 17, pp. 181–204.
13. D. R. Countryman, J. M. Carney, and J. L. Welsh, in Ref. 9, pp. 91–113.
14. G. D. Shook in Ref. 12, Chapt. 70, pp. 799–815.
15. J. A. Bailie, in C. T. Herakovich and Y. M. Tarnopol'skii, eds., A. Kelly and Y. N. Rabotnov, series, eds., *Handbook of Composites*, Vol. 2, Elsevier Publishers B. V., New York, 1989, Chapt. 7, pp. 353–391.
16. M. W. Hyer, in Ref. 15, Chapt. 2, pp. 85–114.
17. M. M. Schwartz, *Composite Materials Handbook*, McGraw–Hill, Inc., New York, 1984, Chapt. 7, Section 7.7.
18. M. F. Ashby and D. R. H. Jones, *Engineering Materials 2: An Introduction to Microstructures, Processing and Design*, Pergamon, Oxford, UK, 1973.
19. R. Nicholls, *Composites Construction Materials Handbook*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1976.
20. A. M. Neville, *Properties of Concrete*, 2d (metric) ed., Pitman Publishing, London, 1973.
21. T. H. Ferrigno, in H. S. Katz and J. V. Milewski, eds., *Handbook of Fillers for Plastics*, Van Nostrand Reinhold Co. Inc., New York, 1984, Chapt. 2, pp. 8–61.
22. L. V. McAdams and J. A. Gannon, in Ref. 6, pp. 258–318.
23. K. Budinski, *Engineering Materials: Properties and Selection*, 2nd ed., Reston Publishing Company Inc., Reston, Va., 1983.
24. D. H. Middleton, ed., *Composite Materials in Aircraft Structures*, Longman Scientific and Technical, London, 1990.
25. F. S. Galasso, *Advanced Fibers and Composites*, Gordon and Breach Publishers, London, 1989.
26. A. R. Bunsell, ed., *Fibre Reinforcements for Composite Materials*, Elsevier Science Publishing Co. Inc., New York, 1988.
27. A. K. Dhingra and H. G. Lauterbach, in Ref. 6, pp. 319–365.
28. J. P. Riggs, in Ref. 6, pp. 20–65.
29. D. M. Riggs, R. J. Shuford, and R. W. Lewis, in Ref. 8, pp. 196–271.

30. M. Lewin and J. Preston, eds., *Handbook of Fiber Science and Technology*, Vol. 3, *High Technology Fibers*, Part A, Marcel Dekker, Inc., New York, 1985.

31. R. Moreton, W. Watt, and W. Johnson, *Nature* **213**, 690 (Feb. 18, 1967).

32. E. Fitzer and M. Heine, in Ref. 26, pp. 73–148.

33. H. H. Yang, in Ref. 26, pp. 249–329.

34. F. E. Wawner, in Ref. 26, pp. 371–425.

35. C. S. Smith, *Design of Marine Structures in Composite Materials*, Elsevier Applied Science Publishers, London, 1990.

36. P. Davies and M. L. Benzeggagh, in K. Friedrich, ed., *Application of Fracture Mechanics to Composite Materials*, Elsevier Science Publishers B. V., New York, 1989, Chapt. 3, pp. 81–112.

37. F.-X. de Charentenay, J. M. Harry, Y. J. Prel, and M. L. Benzeggah, *ASTM STP* **836**, 84–103 (1984).

38. T. T. Serafini, in Ref. 8, pp. 89–114.

39. D. Wilson, in D. Wilson, H. D. Stenzenberger, and P. M. Gergenrother, eds., *Polyimides*, Chapman Hall, New York, 1990, Chapt. 7, pp. 187–226.

40. D. P. Jones, D. C. Leach, and D. R. Moore, *Polymer* **26**, 1385–1393 (Aug. 1985).

41. J. Morton and E. W. Gogwin, *Composite Structures* **13**, 1–19 (1989).

42. G. Dorey, in G. A. O. Davies, ed., *Structural Impact and Crashworthiness*, Elsevier Applied Science Publishers, Barking, UK, 1984, Chapt. 6, pp. 155–192.

43. W. J. Cantwell, P. Davies, P.-Y. Jar, P.-E. Bourban, and H. H. Kausch, in H. Hornfeld, ed., *Plastics–Metals–Ceramics*, SAMPE (European), 1990, pp. 411–427.

44. Aromatic Polymer Composite APC2 Data Sheet 2, I.C.I. Ltd.

45. A. Slobodzinsky, in Ref. 8, pp. 368–390.

46. W. B. Goldsworthy, in Ref. 8, pp. 479–490.

47. B. J. Devlin, M. D. Williams, J. A. Quinn, and A. G. Gibson, *Composites Manufacturing* **3**(4), 203–207 (1991).

48. A. T. Walsh, *Met. Mater.* **7**(10), 606–609.

49. M. G. Phillips, *Compos. Sci. Technol.* **43**, 95–100 (1992).

50. J. C. Halpin and S. W. Tsai, "Effects of Environmental Factors on Composite Materials," *U.S. Air Force Mater. Lab. Rep. AFML-TR 67-423* (1969).

51. S. W. Tsai and H. T. Hahn, *Introduction to Composite Materials*, Technomic, Lancaster, Pa., 1985.

52. S. Kellas and J. Morton, *ALAA J.* **30**(4) 1074–1080 (Apr. 1991).

53. J. Morton, H. Ho, M. Y. Tsai, and G. L. Farley, *J. Compos. Mater.* **26**(5), 708–750 (1992).

54. S. Kellas and J. Morton, *J. Aerospace Eng., I. Mech. E.* **204**, 57–69 (Nov. 1990).

J. Morton
Virginia Polytechnic Institute and State University

W. J. Cantwell
Ecole Polytechnique F d rale de Lausanne

POLYMER-MATRIX

Thermosets,
Thermoplastics,

THERMOSETS

Thermosetting unsaturated polyester resins constitute the most common fiber-reinforced composite matrix today. According to the Committee on Resin Statistics of the Society of Plastics Industry (SPI), 454,000 t of unsaturated polyester were used in fiber-reinforced plastics in 1990. These materials are popular because of their low price, ease of use, and excellent mechanical and chemical resistance properties. Over 227 t of phenolic resins were used in fiber-reinforced plastics in 1990 (1–3). Phenolic resins (qv) are used when their inherent flame retardance, high temperature resistance, or low cost overcome the problems of processing difficulties and lower mechanical properties.

Polyester Resins

Unsaturated polyester resins predominate among fiber-reinforced composite matrices for several reasons. A wide variety of polyesters is available and the composites fabricator must choose the best for a particular application. The choice involves evaluation of fabrication techniques, temperatures at which the resin is to be handled, cure time and temperature desired, and required cured properties (see POLYESTERS, UNSATURATED).

Manufacture. Polyester resins are manufactured by the reaction of dibasic acids with glycols. Common dibasic acids used in polyester production are phthalic anhydride [85-44-9], isophthalic acid [21-91-5], maleic acid [110-16-7], and adipic acid [124-04-9]. The phthalic acids provide stiffness, hardness, and temperature resistance; maleic acid provides the vinyl unsaturation to accommodate free-radical cure; and adipic acid provides flexibility and ductility to the cured resin. Commonly used glycols are propylene glycol [57-55-6], which reduces crystallization tendencies and improves solubility in styrene [100-42-5]; ethylene glycol [107-21-1], which costs the least; and diethylene glycol [111-46-6], which reduces crystallization tendencies. The diacids and glycols are condensed, eliminating water, and then dissolved in a vinyl monomer to a suitable viscosity. Vinyl monomers used are styrene, vinyltoluene [25012-15-4], *p*-methylstyrene [622-97-9], methyl methacrylate [80-62-6], and diallyl phthalate [131-31-9]. The addition of a polymerization inhibitor, such as hydroquinone [123-31-9], *tert*-butylcatechol [98-29-3], or phenothiazine [92-84-2], extends the shelf life of polyester resins.

Resins based on phthalic anhydride are termed *orthophthalic polyesters*, and resins based on isophthalic acid are *isophthalic polyesters*. The most commonly encountered resins are the orthophthalic polyesters, because the isophthalic polyesters, although offering improved mechanical and thermal properties, are higher in cost. Resins based on terephthalic acid [100-21-0] improve upon the property set of isophthalic polyesters, but are very uncommon owing to higher cost.

Polyesters offer low and tailorable viscosities by dilution with vinyl monomers. Low viscosity is important in the fabrication of fiber-reinforced composites to insure good wetting of the fibers and fillers. Poor wetting results in large losses of mechanical properties. Too low a viscosity results in resin-starved areas from gravity drainage. Typically, polyesters are manufactured with styrene concentrations producing resin viscosities of 200 to 1000 mPa·s(–cP), though specialty resins may range from 20 mPa·s(–cP) to 2000 Pa·s(20,000 P). Although styrene monomer dominates the polyester industry, other vinyl monomers are sometimes used in combination with styrene for special purposes. The addition of methyl methacrylate to polyesters improves uv resistance, and the addition of divinyl-benzene [1321-74-0] and diallyl phthalate increases thermal resistance.

Application. Polyesters are cured by free radicals, most commonly produced by the use of peroxides. A wide range of peroxide initiators (qv) are available for use in curing polyesters. Most peroxide initiators are thermally decomposed into free radicals, and the common initiators used at room temperature require the use of a promoter such as dimethylaniline or cobalt octoate.

Because the elevated temperature curing resin systems are thermally activated, they provide a very long time at lower temperatures for fabrication (pot life). Long pot lives make applications such as sheet molding compounds possible; it is the attribute responsible for most of the elevated temperature cure applications. An example of a room temperature curing resin system is 100 parts by weight (pbw) thermosetting polyester resin, 1 pbw methyl ethyl ketone peroxide (90° o), and 0.2 pbw cobalt octoate (6° o). An elevated temperature curing resin system might include 100 pbw thermosetting polyester resin and 1 pbw di-*tert*-butyl peroxide.

Some characteristics restrict the application of polyesters. Polyesters have a limited shelf life, slowly polymerizing over a period of months at room temperature. The shelf life can be extended by cold storage or the addition of polymerization inhibitors by the manufacturer. Styrene emissions occur during the handling and cure of polyesters, and these emissions have increasingly come under government scrutiny. Cured polyesters have relatively low heat resistance with typical *T_g*'s ranging from 60° to 120°C. The higher *T_g* resins are normally brittle with tensile elongation of 1° or less at failure. Polyesters are not resistant to ultraviolet light; the matrix slowly erodes on prolonged exposure. Polyesters have limited resistance to alkaline exposure because of the hydrolyzable ester linkages. Short exposure to strong base causes the resin to soften and become sticky; long exposure dissolves it.

Ease of cure, easy removal of parts from mold surfaces, and wide availability have made polyesters the first choice for many fiber-reinforced composite molders. Sheet molding compound, filament winding, hand lay-up, spray up, and pultrusion are all well adapted to the use of polyesters. Choosing the best polyester resin and processing technique is often a challenge. The polyester must be a type that is well adapted to the processing method and must have the final mechanical properties required by the part application. Table 1 lists the desirable properties for a number of fiber-reinforced composite fabrication methods.

Table 1. Resin Properties Required for Various Fabrication Methods

Process	Viscosity, mPa·s(–cP)	Cure temperature, °C	Thixotropy	Filler, %	Glass, %
SMC BMC ^a	200–2500	150	no	25–50	25–50
hand lay-up	400–800	RT	yes		20–40
filament winding	600–2000	RT-150	yes and no		40–70
pultrusion	400–2000	100–150	no	0–20	60–80
prepreg	50,000+	100–150	no		60–80
spray up	200–1000	RT	yes	0–20	20–40

^a SMC = sheet molding compound; BMC = bulk molding compound.

Most fabricators of fiber-reinforced composites concentrate on only a few of the possible fabrication processes well suited to the parts they produce. Heat cures are used when a large number of parts are made, maximum properties are needed, or the shortest molding cycle time is required. Heat cures require more rigid and expensive molds, along with presses or other equipment to handle these molds. Molding times with heat cures can be as short as 60 s. Room temperature cures can be done on wood, plaster, or plastic molds. These molds are adequate for short runs and are fairly inexpensive. Room temperature cures normally require molding times of an hour to a day, depending on the ambient temperature, thickness, and complexity of the part.

To optimize the resin system for a given process and part, consideration should be given to fillers that can greatly affect the cost and performance of the composite. Because of their low viscosity, fillers can often be added to polyesters. Fillers are often much cheaper than the resin they displace, and they can improve the heat resistance, stiffness, and hardness of the composite. Certain fillers, such as fumed silica, impart thixotropy to the resin, increasing its resistance to drainage.

Health, Safety, and Environment. Manufacturers of fiber-reinforced polyester composites need to be concerned with proper handling of hazardous wastes, emissions of volatile organic compounds, and a host of recent laws and regulations. Of primary concern is worker exposure to, and plant emissions of, styrene. OSHA permissible exposure limits on air contaminants for 1990 placed an 8-h time-weighted average of 50 ppm for styrene. The listing of styrene as a probable carcinogen has led to increased regulation. Styrene's flammability needs to be considered in plant design and resin handling. California has recently further restricted emissions of volatile organic compounds, resulting in additional expenditures for pollution control equipment on the part of many polyester resin users.

Organic peroxides need to be stored separately from the polyester resins and promoters. If a peroxide is contaminated with a promoter, violent decomposition can result. Promoters should always be thoroughly mixed into the resin prior to the addition of the peroxide to prevent violent peroxide decomposition. Peroxides can become unstable if stored for too long or at too high a temperature. Peroxide manufacturers advice for storage and disposal must be strictly followed.

Phenolic Resins

Most processors of fiber-reinforced composites choose a phenol formaldehyde (phenolic) resin because these resins are inherently fire retardant, are highly heat resistant, and are very low in cost. When exposed to flames they give off very little smoke and that smoke is of low immediate toxicity. Phenolic resins (qv) are often not chosen, however, because the resole types have limited shelf stability, both resole and novolac types release volatiles during their condensation cure, formaldehyde [50-00-0] emissions are possible during both handling and cure, and the polymers formed are brittle compared with other thermosetting resins.

Manufacture. Phenolic resins are diverse in structure and functionality. A phenol and less than a stoichiometric amount of formaldehyde are condensed with an acid catalyst resulting in a thermoplastic resin termed a *novolac*. These resins are oligomers that are terminated by phenol groups. They require the addition of a curing agent to effect curing. Novolacs are usually solid resins. Condensing phenol and a stoichiometric amount of formaldehyde with an alkaline catalyst and stopping the reaction while the resin is still a thermoplastic results in a resole. Resoles are terminated by methyl groups. They can be cured by heating or by the addition of a catalyst and are normally produced as liquids. The bulk of the phenolic resins used in fiber-reinforced composites are mixtures of novolacs and curing agents.

Application. Curing agents for novolacs are aldehydes or methylene donors. Paraformaldehyde [30525-89-4], hexamethylenetetramine (HEXA) [100-97-0], formaldehyde, propionaldehyde [123-38-6], glyoxal [107-22-2], and hexamethylmethoxymelamine (HMMM), may be used. HEXA is probably the most common novolac curing agent. On heating HEXA breaks down to ammonia and formaldehyde, forming cross-links between the phenol rings.

Resoles can be cured by the addition of base or by heat alone. Their shelf life is thus limited, which is a significant deterrent to their use in fiber-reinforced composites. Resoles are often used in unreinforced applications in electronics and high moisture areas.

The principal limitation on the use of phenolic resins in fiber-reinforced composites is that they are usually solvent borne and as condensation polymers give off volatile by-products during cure. These volatile by-products consist of water or alcohol that produces property-reducing voids in the cured composite. Phenolic resins are often cured under high pressure to reduce these effects. High pressure curing requires more sophisticated and expensive equipment. Considerable efforts have been expended to develop a room temperature curing phenolic resin for hand lay-up. Some recent success has occurred, making phenolic resins a viable alternative to fire-retardant unsaturated polyester resins. Heat-cured phenolic prepregs have almost completely replaced all other thermosetting resins in aircraft and other transit interiors, where flammability and toxicity of combustion products is a prime consideration.

Health and Safety. Free phenols may be present in phenolic novolacs and resoles. Phenol [108-95-2] is poisonous and caustic, irritating the skin and mucous membranes. Formaldehyde and ammonia [7764-41-7] are often emitted during the cure of novolacs and must be properly vented. Formaldehyde is listed as a human carcinogen; worker exposure and emissions are controlled by OSHA and the EPA.

Epoxy Resins

Currently, epoxy resins (qv) constitute over 90% of the matrix resin material used in advanced composites. The total usage of advanced composites is expected to grow to around 45,500 t by the year 2000, with the total resin usage around 18,000 t in 2000. Epoxy resins are expected to still constitute about 80% of the total matrix-resin-systems market in 2000. The largest share of the remaining market will be divided between bismaleimides and polyimide systems (12 to 15%) and what are classified as other polymers, including thermoplastics and thermoset resins other than epoxies, bismaleimides, cyanate esters, and polyimide systems (see COMPOSITES, POLYMER MATRIX THERMOPLASTICS).

The earliest and still the most widely used matrix resins in high performance composites are the bisphenol A based epoxy resin systems. Full-scale commercial production of epoxy resins began in 1950. Initial significant industrial applications of epoxy resins were in surface coatings and also in potting and encapsulation of electrical components. However, by 1952 epoxy resins were being used in electrical laminates and a filament-wound pressure tank for the F-84 jet fighter. Probably the most widely recognized property of cured epoxy resin systems is their excellent adhesion to a very broad range of substrates and reinforcements. A contributing factor is the low shrinkage exhibited by epoxy resin systems during cure, which results in lower stress levels in the composite than is found in other polymer systems with higher shrinkage. Another factor contributing to the excellent strength of articles produced from epoxy resins is that no by-products are formed during cure. Thus, there are no volatiles liberated that can lead to voids nor are nonvolatiles generated that can act as plasticizers.

Epoxy resins can be generically characterized as a group of commercially available oligomeric materials that contain one or more epoxy (oxirane) groups per molecule. The value of epoxy resins is that they can be processed into a variety of useful products, such as protective coatings, adhesives, and structural components of almost any size and shape, by reaction of the epoxy groups with an appropriate curing agent. The products obtained from epoxy resins that contain more than one epoxy group per molecule are thermosetting polymers.

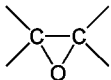


Figure 1 shows the structures of commercial epoxy resins based on a phenolic novolac, an aromatic diacid, an aromatic polyamine (methylenedianiline), and bisphenol A [80-05-7]. Bisphenol A based epoxy resins make up by far the largest portion of the total epoxy resin market. Currently, about 94% of all the epoxy resin sold in the United States is of the bisphenol A or brominated bisphenol A type. This figure does not include epoxidized oils, which are not customarily cured to make thermoset polymers as are standard epoxy resins. Of the remaining 6%, about 3% is made up of specialty resins such as cycloaliphatics, glycidylamines, glycidyl esters, glycidated aliphatic polyethers, and glycidated multifunctional phenols.

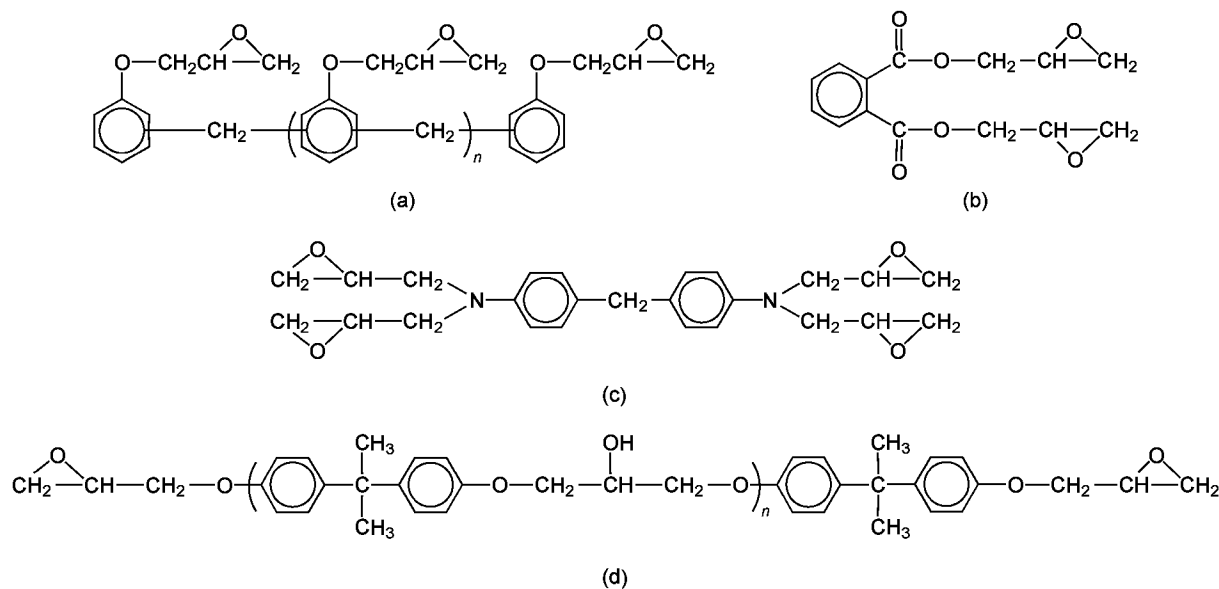


Fig. 1. Structures of commercial epoxy resins: (a) phenolic novolac epoxy resin, (b) glycidated polybasic acid, (c) glycidated polyamine (N,N,N',N'-tetraglycidyl-4,4'-diaminodiphenylmethane [28768-32-3] (TGMDA)), and (d) glycidated bisphenol A.

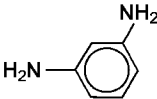
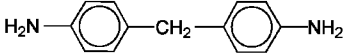
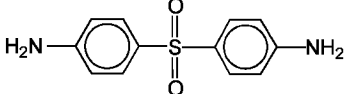
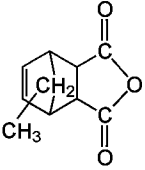
Epoxidized novolacs and high molecular weight epoxy resins, sometimes called phenoxy resins, constitute 2^o and 1^o of the total market, respectively. In commercial products the *n* value ranges from 0 to about 25; the thermoplastic or high molecular weight resins have values of *n* of 200 or more. As *n* increases, the epoxy equivalent weight, or EEW, increases, as does the number of hydroxyl groups. Thus, epoxy resins with lower *n* values are normally used with curing agents that react with the epoxy group, whereas those resins with higher *n* values are cured through the hydroxyl functionality. Typical liquid bisphenol A epoxy resins have *n* values of 0 to 1; resins having values greater than 1 are solid.

Liquid resins such as glycidyl esters and bisphenol A epoxy resins are used mainly in ambient temperature cure coatings, electrical castings, flooring, electrical laminates, and fiber-reinforced composites. These applications require low viscosity materials for good flow and are cured through the epoxide ring. The higher *n* value resins, particularly those above 3000 molecular weight, are normally used in solution and find their greatest application in heat-cured finishes. In these resins the concentration of epoxy groups is low, so they are cured with materials that react with the hydroxyl groups along the backbone.

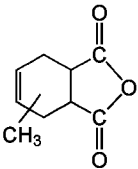
For more demanding uses at higher temperatures, for example, in aircraft and aerospace and certain electrical and electronic applications, multifunctional epoxy resin systems based on epoxy novolac resins and the tetraglycidyl amine of methylenedianiline are used. The tetraglycidyl amine of methylenedianiline is currently the epoxy resin most often used in advance composites. Tetraglycidyl methylenedianiline [28768-32-3] (TGMDA) cured with diaminodiphenyl sulfone [80-08-0] (DDS) was the first system to meet the performance requirements of the aerospace industry and is still used extensively.

Curing Agents. Epoxy resins are reactive intermediates composed of a mixture of oligomeric materials containing one or more epoxy groups per molecule. To convert epoxy resins into useful products they must be cured or cross-linked by chemical reaction into a three-dimensional network. Cross-linking agents, or curing agents, function by reacting with, or causing the reaction of, the epoxide. The two principal classes of curing agents used in epoxy matrix resins for advanced composites are aromatic diamines and anhydrides. Table 2 gives the structures and properties of several of the more commonly used aromatic diamine and anhydride curing agents.

Table 2. Structure and Properties of Aromatic Diamines and Acid Anhydrides Used in Epoxy-Matrix Resins

Name	CAS Registry Number	Molecular weight	Amine equivalent weight	Typical usage, phr	Heat-deflection temperature, ^a °C
<i>Diamines</i>					
<i>m</i> -phenylenediamine	[108-45-2]	108	27	14	150
					
4,4'-methylenedianiline	[101-77-9]	198	50	27	155
					
4,4'-diaminodiphenyl sulfone	[80-08-0]	248	62	18	176
					
<i>Anhydrides</i>					
nadic methyl anhydride	[25134-21-8]	178		80–90	128
					
hexahydrophthalic ^b anhydride	[85-42-7]	154		75–80	130
methyltetrahydrophthalic	[71070-44-3]	166		75–80	130

anhydride



^a Deflection temperature obtained with resin having an epoxide equivalent weight (EEW) of 190.

^b Structure as below; no methyl, no C–C double bond.

Manufacture. Liquid epoxy resins are generally manufactured from bisphenol A and epichlorohydrin [106-89-8]. Typically, the bisphenol A reacts with epichlorohydrin to give a bischlorohydrin, which is dehydrohalogenated with caustic to give the desired epoxy resin. In practice condensation of the bisphenol A with epichlorohydrin and dehydrohalogenation are carried out simultaneously. A consequence of reacting in this manner is that substantial resin is formed before all the phenolic hydroxyl is consumed, leading to attack by the phenol on the resin epoxide instead of on the epichlorohydrin. Glycidylamines such as tetraglycidyl methylenedianiline (TGMDA) are manufactured by a similar process. However, the glycidation of amine is carried out by a two-step procedure because (1) the high reactivity of the basic nitrogen compared with that of the phenolic results in free amine reacting with the epoxide as it is formed, resulting in higher molecular weight or even gelled product, and (2) having all the epichlorohydrin in contact with the amine could lead to a highly exothermic reaction (see EPOXY RESINS).

High Performance Epoxy Resins

Although bisphenol A resins are used extensively in composites, they are limited by the maximum obtainable glass-transition temperature of the bisphenol A systems, which is generally below 180°–190°C. The glass-transition temperature of cross-linked systems is in part related to the cross-link density of the system, and higher *T*_gs can be obtained through additional cross-linking by using epoxy resins of functionality greater than 2. For example, a liquid bisphenol A resin having an epoxy equivalent weight of 189 when cured by **diaminodiphenyl sulfone** has an ultimate glass transition temperature of 175°C. Using a tetrafunctional resin, tetraglycidyl methylenedianiline, with the same curing agent results in *T*_gs greater than 250°C.

Tetraglycidyl methylenedianiline cured with diaminodiphenyl sulfone has been the principal resin system used in high performance composite applications. Although the specific details of most formulated matrix resins are proprietary, formulations of some of the more popular early systems for carbon fiber prepreg based on TGMDA–DDS are given in Table 3. The bisphenol epoxy novolac resin and glycidyl ester are believed to be formulated into the systems to modify the curing characteristics and tack and drape, respectively. The boron trifluoride complex is added as an accelerator. These matrix resin systems offer a combination of ease of melt processing into prepreg as well as tack and drape of the prepreg and out time.

Table 3. Typical Epoxy Resin Prepreg Formulation

Properties	System 1	System 2	System 3
resin components, phr			
tetraglycidyl methylenedianiline ^a	90	100	85
bisphenol epoxy novolac ^b	10		
diglycidyl phthalate ^c			15
diaminodiphenyl sulfone	30	32.0–32.1	34
boron trifluoride complex		1.5	0.5
resin fiber, wt %			
resin	44.6	36.2	40.0
T-300 graphite fiber	58.4		
AS graphite fiber		63.8	60.0

^a ARALDITE MY 720 (trademark of CIBA GEIGY Corp.).

^b Epi-Rez SU-8 (trademark of Rh ne Poulenc).

^c Epi-rez A-100 (trademark of Rh ne Poulenc).

Current TGMDA systems, however, do not meet all the requirements for advanced composites for future aircraft. Specifically, TGMDA systems lack the necessary hot–wet performance and they are too brittle. These shortcomings are inherent in the basic resin, which is a highly polar molecule having a high level of functionality. Since the early 1980s resin suppliers have been working on new systems to meet the requirements for commercial and military aircraft of the 1990s.

The highly polar nature of the TGMDA–DDS system results in high moisture absorption. The plasticization of epoxy matrices by absorbed water and its effect on composite properties have been well documented. As can be seen from Table 4, the TGMDA system can absorb as much as 6.5% (by weight) water (4). This absorbed water results in a dramatic drop in both the glass transition temperature and hot–wet flexural modulus (4–6).

Table 4. Effect of Absorbed Water on the Mechanical Properties of the TGMDA–DDS System^a

Properties	Stoichiometry of DDS	
	54% _w	100% _w
components, phr		
tetraglycidyl methylenedianiline	100	100
bisphenol A epoxy novolac ^b	8.2	8.2
4,4-diaminodiphenyl sulfone	28.0	51.9
glass-transition temperature, °C		
dry	242	
wet	174	
flexural properties, RT dry		
strength, MPa ^c	117	131
modulus, GPa ^d	4.0	3.0
elongation, % _w	3.5	4.7
flexural properties, hot wet ^e		
strength, MPa ^c	84	74

modulus, GPa ^d	3.2	24
elongation, % ^o	3.4	3.7
moisture gain, wt % ^{o f}	4.7	5.8

^a Cured 2 h at 150° and 4 h at 200°.

^b Epi-Rez SU-8 (trademark of Rhϕne Poulenc).

^c To convert MPa to psi, multiply by 145.

^d To convert GPa to psi, multiply by 145,000.

^e Tested in water at 93°C after 2 weeks' immersion at 93°C.

^f Conditioned 2 weeks at 93°C.

New products designed to overcome these shortcomings are beginning to appear. The Dow Chemical Company has introduced a glycidyl ether based on a hydrocarbon epoxy novolac (7) under the trade name TACTIX 556. The structure of the resin is given in Figure 2, and the properties of neat resin castings are given in Table 5. This multifunctional epoxy resin has a glass transition temperature of over 223°C when cured with DDS and outstanding moisture resistance and thermal oxidative resistance. Moisture absorption in boiling water is reported to be about 2.4 wt % for TACTIX 556 cured with DDS. Under the same conditions the corresponding TGMDA system has about a 4.5 wt % equilibrium water uptake.

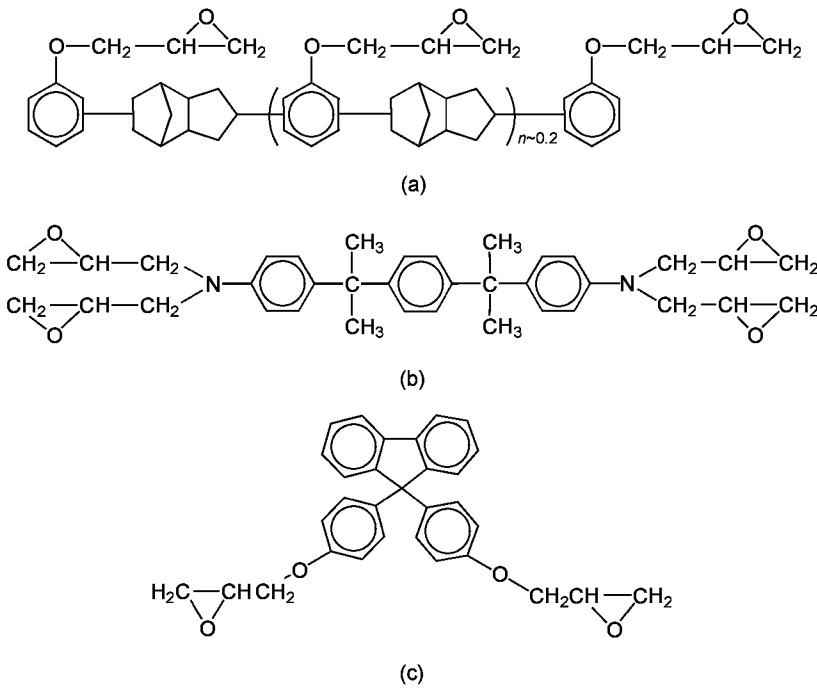


Fig. 2. Epoxy matrix resins for advanced composites: (a) TACTIX 558 (Courtesy of The Dow Chemical Company), (b) EPON HPT Resin 1071 (Courtesy of Shell Chemical Co.), and (c) EPON HPT Resin 1079.

Table 5. Mechanical Properties of High Performance Epoxy Resins^a in Unreinforced Resin Castings

Properties	TGMDA	EPON HPT ^b 1071 ^c	EPON HPT ^b 1079 ^c	TACTIX ^{d,e}
T_g , °C		249	279	556
flexural properties RT ^o dry				
strength, MPa ^f	138	117	124	136
modulus, GPa ^g	3.9	3.9	3.3	3.1
elongation, % ^o	5.0	3.7	4.7	
flexural properties hot ^o wet ^h				
strength, MPa ^f	76	90	90	
modulus, GPa ^g	2.5	3.4	3.0	3.0 ^h
elongation, % ^o	4.7	3.6	4.1	
moisture gain, wt % ^{o i}	5.7	3.6	2.8	2.4 ^j

^a Resin cured with 100% diaminodiphenylsufone.

^b Trademark of Shell Chemical Co.; see Fig. 2.

^c Cured 2 h at 150° C and 4 h at 200°C.

^d Trademark of The Dow Chemical Company.

^e Cured 3 h at 177°C and 2 h 232°C.

^f To convert MPA to psi, multiply by 145.

^g To convert GPa to psi, multiply by 145,000.

^h Tested in water at 93°C after 2 weeks' immersion at 93°C.

ⁱ Specimens conditioned 200 h in boiling water.

^j Conditioned 2 weeks in 93°C water.

^k Conditioned 200 h in boiling water.

Shell Chemical Company has introduced a new tetraglycidyl amine and a new stiff backbone diglycidyl ether under the respective trade names EPON HPT Resin 1071 [103490-06-8] and EPON HPT Resin 1079 [47758-37-2] (8,9), which have superior hot-wet performance compared with TGMDA. These two materials are chemically *N,N,N',N'*-tetraglycidyl- α,α' -bis(4-aminophenyl)-*p*-diisopropylbenzene and diglycidyl-9,9-bis(4-hydroxyphenyl) fluorene. The structures of these two resins are given in Figure 2. As can be seen from Table 5, the moisture gain (after 2 weeks' immersion in water at 93°C) for EPON HPT Resin 1071 and EPON HPT Resin 1079 cured with DDS is 3.7 and 2.8 wt %, compared with 5.7 wt % for the corresponding TGMDA system. This results in a much improved hot-wet performance for these two new resins over TGMDA. Specifically, under hot-wet conditions, under water at 93°C after 2 weeks' immersion at 93°C, the TGMDA system retained only 65% of its dry room temperature modulus, whereas the EPON HPT Resin 1071 and EPON HPT Resin 1079 retained 85 and 91%, respectively, of their dry room temperature modulus under hot-wet conditions.

Although changes in the structure of highly functional epoxy resins have resulted in improved hot-wet performance, brittleness is an inherent property of highly cross-linked systems. Improvements in ductility, fracture resistance, and impact strength, however, can be achieved without substantially degrading the thermal and mechanical properties of the epoxy matrix. For example, carboxyl terminated butadiene-acrylonitrile copolymer, which is initially soluble in conventional bisphenol A epoxy resins, forms a second phase of dispersed particles on cure, resulting in an epoxy resin system with improved toughness (10). The degree to which any epoxy resin system can be toughened depends on many factors, including the type of modifier, heterophase size distribution, interfacial adhesion between the heterophase and the matrix, combination of resin and curing agent, cure conditions, and molecular weight between cross-links in the matrix. A thorough and critical review of all the variables affecting the ability to toughen epoxy resin systems is beyond the scope of this discussion (see Ref. 11).

The use of elastomeric modifiers for toughening thermoset resins generally results in lowering the glass transition temperature, modulus, and strength of the modified system. More recently, ductile engineering thermoplastics and functional thermoplastic oligomers have been used as modifiers for epoxy matrix resins and other thermosets (12).

Applications. Epoxy resins constitute over 90% of the matrix resin material used in advanced composites. In addition, epoxy resins are used in all the various fabrication processes that convert resins and reinforcements into composite articles. Liquid resins in combination, mainly, with amines and anhydride are used for filament winding, resin transfer molding, and pultrusion. Parts for aircraft, rocket cases, pipes, rods, tennis rackets, ski poles, golf club shafts, and fishing poles are made by one of these processes with an epoxy resin system.

Bismaleimides

Bismaleimides (BMI) are a relatively young class of thermosetting polymers that are gaining acceptance by the industry because they combine a number of unique features including excellent physical property retention at elevated temperatures and in wet environments, almost constant electrical properties over a wide range of temperatures, and nonflammability properties. Bismaleimides have become a leading class of thermosetting polyimides. Their excellent processibility and balance of thermal and mechanical properties have made them popular in advanced composites and electronics.

Bismaleimides are best defined as low molecular weight, at least difunctional monomers or prepolymers, or mixtures thereof, that carry maleimide terminations (Fig. 3). Such maleimide end groups can undergo homopolymerization and a wide range of copolymerizations to form a highly cross-linked network. These cure reactions can be effected by the application of heat and, if required, in the presence of a suitable catalyst. The first patent for cross-linked resins obtained through the homopolymerization or copolymerization of BMI was granted to Rhône Poulenc, France, in 1968 (13). Shortly after, a series of patents was issued on poly(amino bismaleimides) (14), which are synthesized from bismaleimide and aromatic diamines.

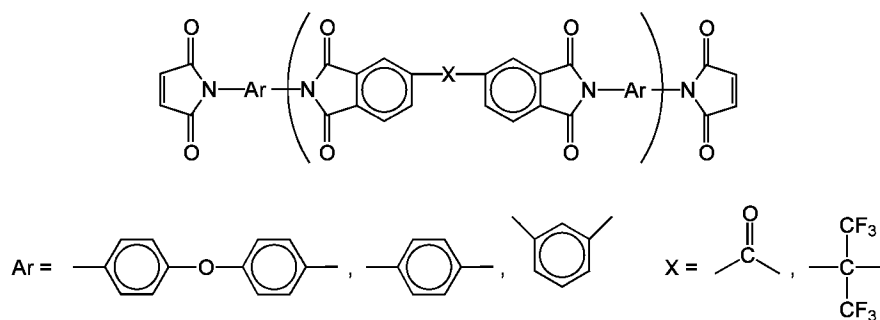


Fig. 3. Structures of bismaleimide resins.

A number of BMI resins based on this chemistry became commercially available through Rhône Poulenc for application in printed circuit boards and molding compounds and Rhône Poulenc recognized the potential of bismaleimides as building blocks for temperature-resistant thermoset systems. The basic chemistry, however, was not new, because the Michael addition reaction had been employed by Du Pont to obtain elastomeric reaction products from bismaleimides and liquid polymeric organic diamines (15).

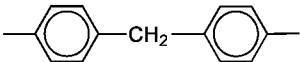
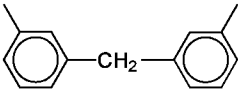
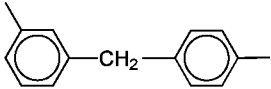
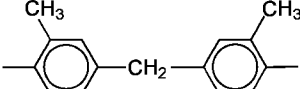
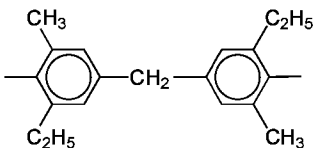
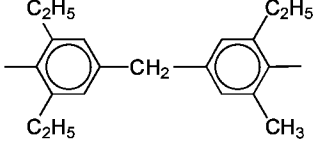
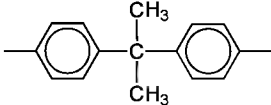
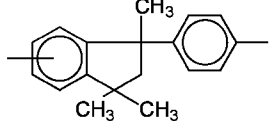
Following the initial patent disclosures, research organizations worldwide have worked in this new area as evidenced by approximately 500 patent applications in the following 20 years. Numerous chemical modifications can be carried out (16) and mixtures formulated to improve rheology and cured resin properties. Melt processible bismaleimide systems specifically designed for fiber reinforced applications are disclosed in a U.S. patent granted to Technochemie (17). The approach of blending selected bismaleimide monomers and further formulating them with reactive diluents or comonomers provided formulated bismaleimide resins that are equivalent in processing to the conventional 177°C service epoxy resins. The breakthrough with respect to application as tacky autoclavable prepreg was achieved by US-Polymeric (now British Petroleum Chemicals) with their V378A BMI system (18). This system was used to build the delta wings of the General Dynamics F16XL demonstrator combat aircraft (19) and found application for parts of the McDonnell Douglas AV8-B vertical takeoff aircraft (20).

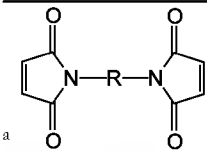
The principal concern with BMI resins has been their inherent brittleness owing to their high cross-link density. Ciba Geigy, however, demonstrated that BMI-*o,o'*-diallylbisphenol A copolymers are tougher than TGMDA-DDS epoxy resins (21). The *o,o'*-diallylbisphenol A-BMI copolymers were patented in Switzerland in 1975, but the significance of this invention, that is, their toughness, was not recognized before toughness became an issue to the aerospace industry. Experience has taught that a useful BMI-resin system comprises both a bismaleimide part (the BMI resin) and a comonomer part (reactive diluent).

Building Blocks and Systems. A standard synthesis of *N,N'*-arylene bismaleimide involves the chemical dehydration of *N,N'*-arylene bismaleamic acid with acetic anhydride [108-24-7] and sodium acetate [127-09-31] as a catalyst at temperatures below 80°C (22). The yield of pure recrystallized BMI is usually 65–75%. Various by-products, such as isoimides and acetanilides, are responsible for the relatively low yield of pure BMI (23). Almost every aromatic amine (diamine) can be converted to the corresponding maleimide (bismaleimide). The most widely used building block is 4,4'-bismaleimidodiphenylmethane [13676-54-5], because the precursor diamine is readily available and cheap. Bismaleimides based on aromatic diamines are crystalline substances with high melting points. Table 6 gives the chemical structures and melting points of bismaleimides based on substituted and isomeric diphenylmethanes. Of interest to resin formulators, whose aim is to achieve unique properties in resins and adhesives, are the melting points and polymerization characteristics, which are used to tailor the flow and cure properties of their products. For reasons of processibility BMIs with low melting

points are the preferred building blocks.

Table 6. Structures and Properties of Bismaleimides

R ^a	Mp, °C	T _{max} ^b , °C	ΔH _{pol} ^c , J g ^d
	155–157	235	198
	195–196		
	164–165		
	210–212		
	150–154	298	187
	149–151	328	206
	235	290	216
	90–100	203	89



^a From differential scanning calorimetry (dsc) data from Technochemie GmbH–Verfahrenstechnik. Heating rate = 10° C min;
T_{max} = cure exotherm peak maximum.
^c ΔH_{pol} = heat of polymerization from dsc data.
^d To convert J to cal, divide by 4.184.

The order of melting transitions for the diphenylmethane bismaleimides is surprising (Table 6) because the 4,4'-MDA-BMI melts lower (155–157°C) than the 3,3'-MDA-BMI. The tetra alkyl-4,4-MDA-BMI also melts around 150°C. The BMI based on the isomeric diphenylindane diamine (24) is very low melting because it is a complex blend of various isomers.

The most important property of a bismaleimide is its ability to undergo a temperature-induced polymerization. The maleimide double bond is highly activated owing to the adjacent carbonyl groups of the imide ring; therefore, heating the BMI above its melting point effects polymerization. The cure exotherm peak temperature taken from a dynamic dsc scan is used to characterize the reactivity of individual BMIs. The reactivity is influenced by the chemical nature of the residue between the maleimide terminations and by the molar mass between the reactive maleimide groups. Normally, electron-donating groups, such as alkyl groups, reduce the BMI reactivity, provided that they are present in the maleimide rings that carry the maleimide groups. On the other hand, electron-attracting groups such as SO₂, carbonyl, and so on, have the opposite effect. Increasing the molecular weight between cross-links (*M*_c) generally provides a more latent system owing to steric effects.

The BMI-building blocks primarily used in commercial bismaleimide resins are 4,4'-bismaleimidodiphenylmethane, 2,4-bismaleimidotoluene [6422-83-9], 1,3-bismaleimidobenzene [3006-93-7], and, sometimes, aliphatic BMIs based on *n*-alkanes. However, because of toxicity problems associated with MDA (4,4'-diaminodiphenylmethane) and other diamines with only one or two aromatic rings, polyaromatic diamines and BMIs based on them are

of increasing interest (Fig. 4) (see AMINES, AROMATIC—METHYLENEDIANILINE).

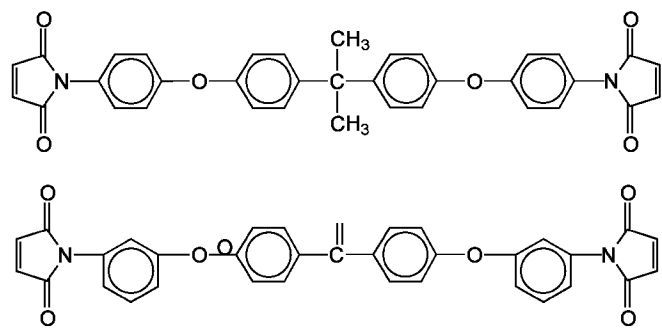


Fig. 4. Bismaleimides of polyaromatic diamines.

In the past, it was almost impossible to introduce novel BMI building blocks with increased molecular weight because of the processing problem associated with high melting points and high melt viscosities. With new processing techniques, such as powder prepregging and blending with liquid reactive diluents such as *o,o'*-diallylbisphenol A [1745-89-7], such limitations could be partially overcome. It is questionable, however, whether such expensive building blocks will find application in commercial bismaleimide systems.

Besides low molecular weight building blocks, long-chain maleimide-terminated oligomers have been synthesized for molding, adhesive, and composite applications. The key step is the preparation of an amine-terminated intermediate needed to introduce the maleimide group. For example, long-chain thermosetting arylimides synthesized from long-chain **sulfone–ether diamines** and maleic anhydride have been described. The resulting bisamic acid was cyclodehydrated in the usual way with acetic anhydride–sodium acetate (25).

The cured arylimides are heat stable systems possessing high glass transition temperatures (T_g) and moderate to high modulus plateaus above their T_g , depending on the molecular weight between the (maleimide) cross-links. The sulfone–ether backbone (see Fig. 5) alleviates problems associated with solubility of the uncured material. High solids solutions can be prepared in common organic solvents.

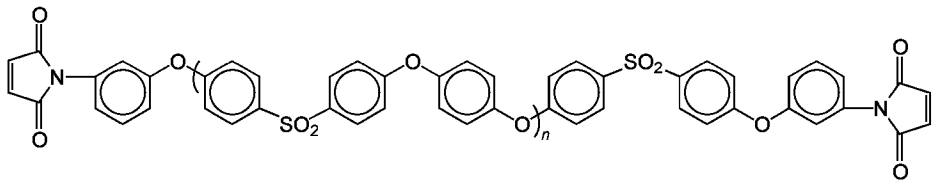
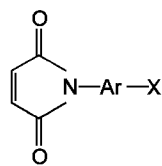


Fig. 5. Structure of a maleimide-terminated sulfone–ether prepolymer.

Other backbone structures that have generated a great interest are the polyether ketones. An attempt was made to synthesize amino-terminated arylene ether ketones, which were subsequently converted into the corresponding maleimide-terminated oligomers (26). The aim of this approach was to obtain tough, solvent-resistant, high temperature thermosets.

The common synthetic route to bismaleimides or maleimide functionalized oligomers is the condensation of diamines or amino-terminated oligomers with maleic anhydride. Another possibility is the use of an AB-type monomer of the following general formula to build the polymaleimide, where X represents a functional group that can be employed in condensation reactions.



The maleimide is prebuilt into the molecule in a separate step. Maleimidobenzoic acid [17075-07-7] or its acid halide was used to synthesize maleimide-terminated polyamides (16,17) or polyesters (27) from amino- or hydroxy-terminated polyamides and polyesters, respectively. The literature on bismaleimide prepolymers and bismaleimide building blocks is quite extensive (28), but only a limited number of BMI building blocks have been used for commercial resin formulations.

Bismaleimide Resin Concepts. The bismaleimide building block is not the final resin product. Although building blocks usually make up 50–75% by weight of the resin, other ingredients, such as comonomers, reactive diluents, processing additives, elastomers, and catalysts, are combined with BMI so as to obtain a product suitable for the application considered. The application areas for bismaleimide resins are reinforced composites for printed circuit boards (with glass fabric), structural laminates (with glass, carbon, and aramid fibers), and moldings (with short fibers and particulate fillers).

In order to fulfill the processing requirements, bismaleimide building blocks have to be formulated into products that enable their use as highly concentrated solutions, powders, or hot melts.

An early attempt in BMI modification with respect to achieving a meltable, noncrystalline resin was made by blending a complex combination of bismaleimides and crystallization inhibitors (29). The resin (COMPIMIDE 353 [51569-11-0], now Shell Chemical Company Technochemie), although it is not a formulated product, shows a very low melting transition and is a low viscosity fluid at 110°C. (Table 7). Such noncrystalline low melting BMI resins can be further formulated with a wide range of vinyl (or allyl, etc), compounds to provide prepreg resins that are tacky at room temperature and suitable for low pressure autoclave molding.

Table 7. COMPIMIDE 353 Resin Data

Property	Test method ^a	Value	Comment
physical form			resolidified melt
appearance			yellow–red brown transparent mass
gel time, min	DIN 16945	35–65	
viscosity mPa·s(cP)	DIN 16945	400–1400	
dsc			

T_g , °C	TC-TM 14	193 ± 10	heating rate 20°C min
T_{max} , °C		275 ± 15	
polymerization energy, J g ^b	TC-TM 14	Δ <i>H</i> 220 ± 40	
composition	TC-TM 25	comparison with standard	hplc

^a DIN German industry norm; TC-TM Technochemie test method.

^b To convert J to cal, divide by 4.184.

The divinylbenzene approach to modify BMI into a hot-melt prepreg resin was followed by US-Polymeric (18). The resin system, known in the industry under the trade name V-378A, was the first BMI formulation used in a primary aircraft application (30). Although this resin is relatively brittle in comparison with epoxy resins, it provides outstanding high temperature mechanical properties in both dry and wet environments. V 378 A prepreg can be cured under standard low autoclave pressure of 0.7 MPa (7 bars) and temperatures similar to current 177°C curing epoxy systems. However, postcure temperatures of 245–260°C are required to maximize elevated temperature strength retention.

Michael Additions. The reaction of a bismaleimide with a functional nucleophile (diamine, bisthiol, etc) via the Michael addition reaction converts a BMI building block into a polymer. The nonstoichiometric reaction of an aromatic diamine with a bismaleimide was used by Rhône Poulenc to synthesize polyaminobismaleimides as shown in Figure 6 (31).

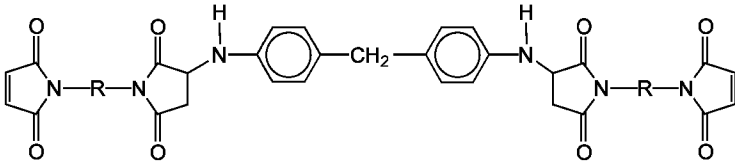
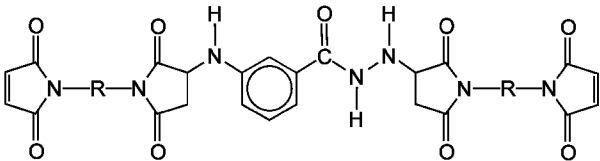


Fig. 6. Michael addition product of bismaleimide.

The reaction product of 4,4'-bismaleimidodiphenylmethane and 4,4'-diaminophenylmethane, known as Kerimide 601 [9063-71-2], is prepolymerized to such an extent that the resulting prepolymer is soluble in aprotic solvents such as *N*-methylpyrrolidinone [872-50-4], dimethylformamide [68-12-2], and the like, and therefore can be processed via solution techniques to prepreg. Kerimide 601 is mainly used in glass fabric laminates for electrical applications and became the industry standard for polyimide-based printed circuit boards (32).

Another approach to processible bismaleimide resins via a Michael addition chain extension, is the reaction of bismaleimide, or a low melting mixture of bismaleimides, with aminobenzoic hydrazide to provide a resin that is soluble in various solvents, such as acetone [67-64-1], methylene chloride [75-09-2], and dimethylformamide [68-12-2] (33). The idealized chemical structure for a 2:1 BMI-aminobenzoic hydrazide resin is as follows:



Two resin systems based on this chemical concept are commercially available from Shell Chemical Company Technochemie under the COMPIMIDE trademark: COMPIMIDE 183 (34) [98725-11-2], for use in printed circuit boards, and COMPIMIDE 796 [106856-59-1], as a resin for low pressure autoclave molding (35). Typical properties of COMPIMIDE 183 glass fabric-PCB laminates are provided in Table 8. COMPIMIDE 183 offers a combination of advantageous properties, such as a high glass transition temperature, low expansion coefficient, and flame resistance without bromine compound additives.

Table 8. Properties of BMI^a Printed Circuit Laminates^b

Properties	Value
dielectric constant (1 MHz)	4.6–4.7
dielectric loss constant (1 MHz)	0.01
dielectric strength, V μm	29.5
volume resistivity, Ω·cm	10 ¹⁵
water absorption, % °	<1
peel strength of copper foil, ^d N mm	1.2–1.4
heat stability at 287°C, s	>60
thermal expansion coefficient, 10 °C	
x,y direction	14–16
z direction	36–38
flammability in comparison with UL 94 ^e	V ⁰ –V ¹

^a COMPIMIDE 183.

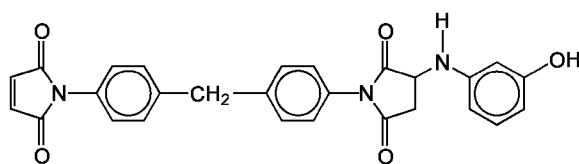
^b Properties measured using 1–1.5 mm laminates; glass fabric U.S. 2116, resin content 45% by weight.

^c After postcuring at 210°C.

^d Gould TC polyimide grade 35-μm thickness; to convert N mm to ppi, divide by 0.175.

^e V⁰ rating for 3-mm laminates; V¹ rating for 1.5-mm laminates.

The Michael addition reaction has attracted many researchers as a route to convert high melting BMI building blocks into resins with improved processibility as compared with the BMI precursors. Heat-resistant resin compositions are prepared from BMI and *para*- or *meta*-aminophenol (38). The idealized structure of such a BMI-*m*-aminophenol adduct follows.



BMI-*m*-aminophenol adducts can further react with epoxy resin and subsequently be cured with imidazole catalyst. 1:2 BMI-aminophenol adducts have been used as curing agents for epoxy resin (37).

An interesting approach to thermosetting acetylene-terminated polyimides via the Michael addition reaction has appeared (38). Acetylene-terminated aspartimides are readily prepared in high yield via two routes, shown in Figure 7.

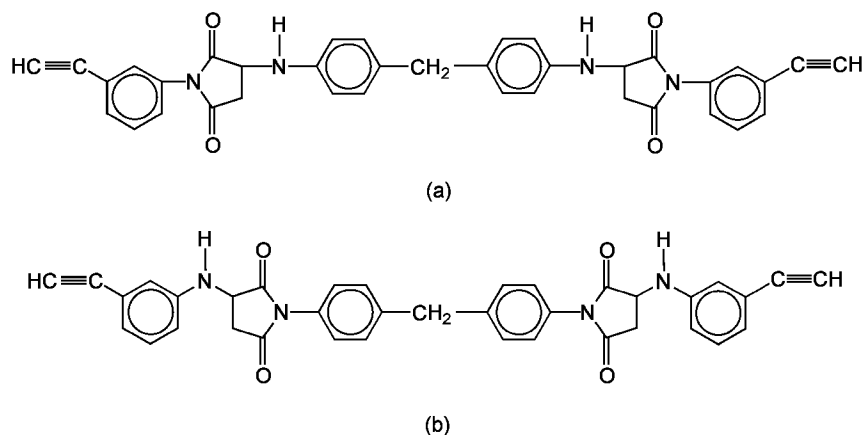
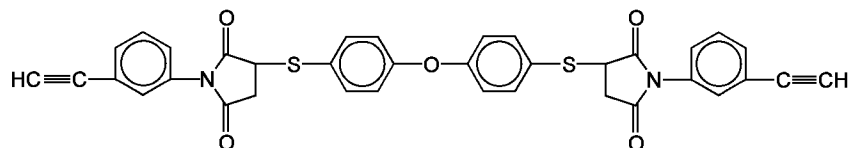


Fig. 7. Acetylene-terminated aspartimides. (a) Reaction of aromatic diamine with *N*-(3-ethynylphenyl)maleimide [105280-01-1] in a 1:2 molar ratio yields a prepolymer of this general formula. (b) Bismaleimide reacts with 3-ethynylaniline [54060-30-9] in a 1:2 molar ratio to yield a prepolymer of this general formula.

The Michael addition of nucleophiles to the carbon-carbon double bond of maleimide has been exploited in the synthesis of a variety of linear polymers through reaction of bismaleimide with bisthiols (39). This method has been used to synthesize ethynyl-terminated imidothioether from the reaction of 4,4'-dimercaptodiphenyl ether [17527-79-6] and *N*-(3-ethynylphenyl)maleimide (40). The chemical structure of this Michael addition imide thermoset is as follows:



The Michael addition reaction of amines and thiols with bismaleimides or functionalized monomaleimides is a versatile tool in the synthesis of chain-extended maleimide-terminated prepolymers. These prepolymers generally are soluble in organic solvents from which they can be processed to prepreg and molded to high quality, void-free laminates.

Bismaleimide Resins via ENE Reaction. The copolymerization of a BMI with *o,o'*-diallylbisphenol A [1745-89-7] (DABA) is a resin concept that has been widely accepted by the industry because BMI-DABA blends are tacky solids at room temperature and therefore provide all the desired properties in prepreps, such as drape and tack, similar to epoxies. Crystalline BMI can easily be blended with DABA, which is a high viscosity fluid at room temperature. Upon heating BMI-DABA blends copolymerize via complex ENE and Diels-Alder reactions as outlined in Figure 8.

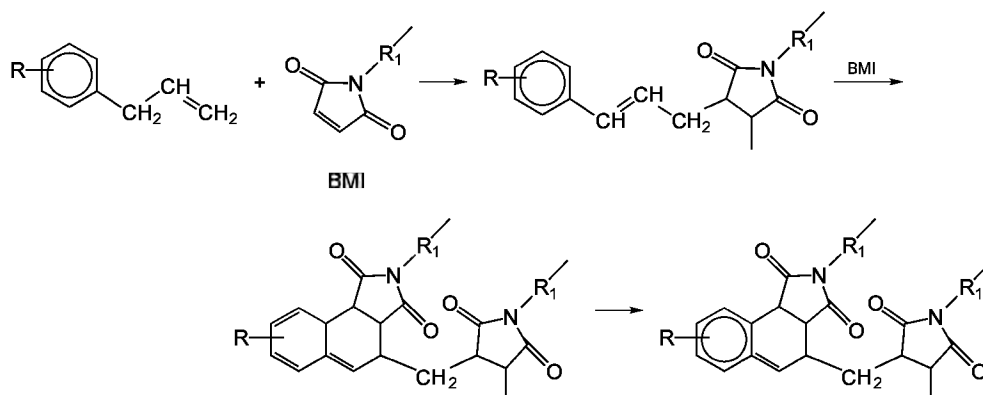
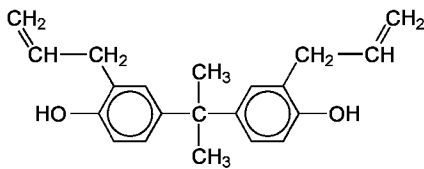


Fig. 8. Polymerization of bismaleimides with *o,o'*-diallylbisphenol A via complex ENE-Diels-Alder reactions.

Reportedly, *o,o'*-diallylbisphenol A is an attractive comonomer for bismaleimides because the corresponding copolymer is tough and temperature resistant (41). Toughness, however, is a function of the BMI-diallylbisphenol A ratio employed. In one study optimized toughness properties were achieved when BMI and diallylbisphenol were employed at a close to 2:1 molar ratio (42). In Table 9, the mechanical properties of BMI-bis(3-allyl-4-hydroxyphenyl)-*p*-diisopropylbenzene resins are provided, showing optimized properties for the 60/40 BMI-diallylbisphenol composition. The *o,o'*-diallylbisphenol A is commercially available under the trademark Matrimide 5292. Another bisallylphenyl compound is available from Shell Chemical Company Technochemie under the trademark COMPIMIDE 121.



o,o'-diallylbisphenol A

Both materials meet the processing requirements of the industry because they are honeylike fluids at room temperature and low viscosity liquids between approximately 70° and 90°C and thus act as reactive diluents and tougheners at the same time.

Table 9. Mechanical Properties of BMI–bis(3-allyl-4-hydroxyphenyl)-*p*-diisopropylbenzene Blends^a

Property ^b	Composition, % BHPDB			
	20	30	35	40
flexural properties ^c				
room temperature				
strength, MPa ^d	117	133	140	160
modulus, GPa ^e	4.69	4.48	4.09	4.46
elongation, %	2.48	3.00	3.46	3.78
177°C				
strength, MPa ^d	90	107	117	133
modulus, GPa ^e	3.75	3.59	3.39	3.47
elongation, %	2.42	2.76	3.32	4.56
250°C				
strength, MPa ^d	79	84	90	75
modulus, GPa ^e	2.86	2.91	2.89	2.79
elongation, %	2.90	3.22	4.44	>5
fracture energy, <i>G</i> _{IC} , J m ^{2f}	30	60	98	217
moisture gain ^g , %	3.33	3.02	3.04	3.18

^a BMI is a eutectic bismaleimide blend; the other component is abbreviated BHPDB.

^b Cure cycle: 3 h at 160°C, 4 h at 210°C, and 5 h at 240°C.

^c Flexural properties were determined dry at various temperatures.

^d To convert MPa to psi, multiply by 145.

^e To convert GPa to psi, multiply by 145,000.

^f To convert J m² to ft lbf in.², divide by 2100.

^g 1000 h, 94% RH, 70°C.

The synthesis of bis[3-(2-allylphenoxy)phthalimides] and their copolymer properties with BMI have been reported (43). These allylphenoxyimide–BMI copolymers provide toughness and temperature resistance when used in carbon fiber laminates (44).

Diels-Alder Copolymers. The Diels-Alder reaction can also be employed to obtain thermosetting polyimides. If bismaleimide (the bisdienophile) and the bisdiene react nonstoichiometrically, with bismaleimide in excess, a prepolymer carrying maleimide terminations is formed as an intermediate, which can then be cross-linked to yield a temperature-resistant network.

Styrene can react with bismaleimide via a complex Diels-Alder-ENE route in a 1:2 stoichiometric ratio as outlined in Figure 9. Other vinylbenzene compounds, such as propenylphenoxydiphenyl sulfone [106818-12-6] (45) and bis(*o*-propenylphenoxy)benzophenone [109423-33-8] (46), react in a similar way with bismaleimide. Their synthesis involves a straightforward nucleophilic halogen displacement reaction. *o*-Allylphenol [1745-81-9] reacts with 4,4'-dichlorodiphenyl sulfone [80-07-9] or 4,4'-difluorobenzophenone [345-92-6] at 160°C in *N*-methylpyrrolidinone [872-30-4] as a solvent in the presence of potassium carbonate [584-08-7] as a catalyst. The alkaline reaction conditions are responsible for the *o*-allyl → *o*-propenyl isomerization. A wide variety of structurally similar bis(propenylphenoxy) compounds are possible by simply using *p*-substituted allylphenols (eugenol) or isomeric dihalosulfones or dihalobenzophenones in the synthesis. The bis(*o*-propenylphenoxy) compounds are low melting, low viscosity materials that can easily be melt blended with bismaleimide and then cured at temperatures of 170–230°C. Table 10 provides the mechanical properties of a series of cured resins. These copolymer resins are attractive because they show improved toughness and reduced moisture absorption in comparison with the unmodified BMI, and also have a lowered glass-transition temperature. Bis(*o*-propenylphenoxy)benzophenone resins are commercially available from Shell Chemical Company Technochemie under the trademark COMPIMIDE 123 [109423-33-8].

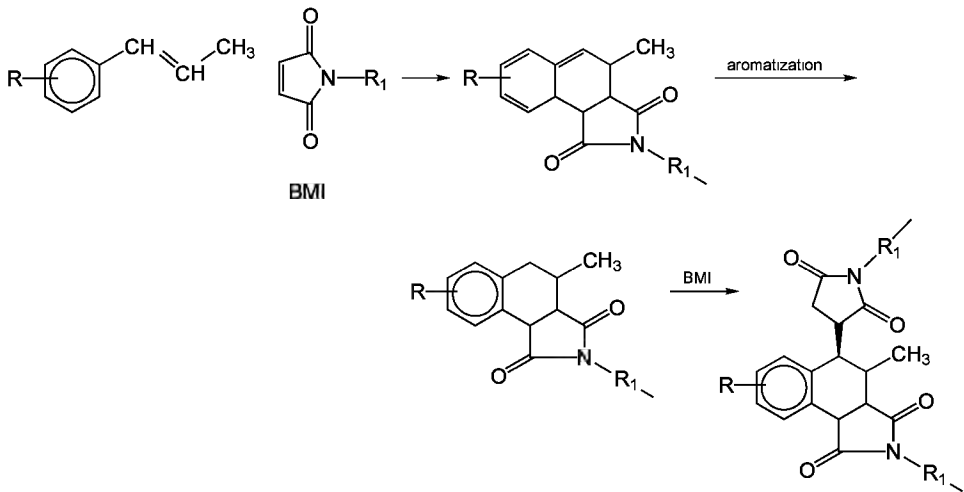


Fig. 9. Reaction of β -methylstyrene with a bismaleimide via a complex Diels-Alder-ENE route.

Table 10. Mechanical Properties of Bismaleimide Diels-Alder Resins^a

BMI, ^a ° o	Composition toughener, ^b ° o	Flexural strength		Flexural modulus,		Flexural		G_e , J m ^{-2e}	T_g^f , °C	Water absorption, ° o
		MPa ^c	250°C	23°C	250°C	23°C	250°C			
100		76	31	4.64	3.03	1.7	1.03	63	>300	4.30
82	18 ^g	98	70	3.99	2.93	2.49	2.37	185	285	4.00
60	40 ^g	114	73	3.58	2.15	3.20	4.50	267	256	2.90
80	20 ^h	87	56	3.85	2.82	2.3	2.0	234	300	3.74
60	40 ^h	128	83	3.49	2.38	3.9	>5	378	277	3.63
80	20 ⁱ	106	65	3.96	2.66	2.34	2.52	191	275(266)	3.66
60	40 ⁱ	132	56	3.70	1.71	3.75	4.86	439	261(249)	2.59
80	20 ^j	114	78	4.17	2.47	2.87	3.73	247	273(252)	3.46
60	40 ^j	122	81	3.59	2.44	3.44	4.52	466	265(260)	2.90

^a BMI COMPIMIDE 796.
^c To connect MPa to psi, multiply by 145.
^d To convert GPa to psi, multiply by 145,000.
^b Diphenyl sulfones or benzophenones as indicated.
^e To convert J m² to ftlb/in.², divide by 2100.
^f By dma analysis. Values in parens by tma analysis.
^g 4,4′Bis(*o*-propenylphenoxy)diphenyl sulfone.
^h 4,4′Bis(*o*-methoxy-*p*-propenylphenoxy)diphenyl sulfone.
ⁱ 4,4′Bis(*o*-propenylphenoxy)benzophenone.
^j 4,4′Bis(*o*-methoxy-*p*-propenylphenoxy)benzophenone.

Propenylphenoxy compounds have attracted much research. BMI–propenylphenoxy copolymer properties can be tailored through modification of the backbone chemistry of the propenylphenoxy comonomer. Epoxy resins may react with propenylphenol (47,48) to provide functionalized epoxies that may be low or high molecular weight, liquid or solid, depending on the epoxy resin employed. Bis[3-(2-propenylphenoxy)phthalimides] have been synthesized from bis(3-nitrophthalimides) and *o*-propenylphenol sodium involving a nucleophilic nitro displacement reaction (49). They copolymerize with bismaleimide via Diels-Alder and provide temperature-resistant networks.

Under appropriate thermal conditions, the strained four-membered ring of benzocyclobutene undergoes electrocyclic ring opening to generate, *in situ*, *o*-chinodimethane, which, in the presence of bismaleimide, reacts via a Diels-Alder reaction (50). The chemistry of the BMI–benzocyclobutene copolymerization is provided in Figure 10. Both bismaleimide and bisbenzocyclobutene can undergo thermal homopolymerization. Therefore, blends of bismaleimide (BMI) and bisbenzocyclobutene (BCB) with either BMI or BCB in excess provide thermosetting polyimides with either maleimide or benzocyclobutene terminations. Certain bisbenzocyclobutene–BMI (COMPIMIDE 353) blends form compatible mixtures in a wide range of ratios, which after cure, show remarkable thermal oxidative stability and high glass transition temperatures (51). Interestingly, it can be demonstrated that the thermal oxidative stability of nonstoichiometric BMI–BCBs with excess BMI is far superior to the BMI homopolymer. BMI–BCB systems, however, are not yet commercial. Another not yet commercial family of dienes, the bis(3,4-dimethylene-pyrrolidines), has emerged (40).

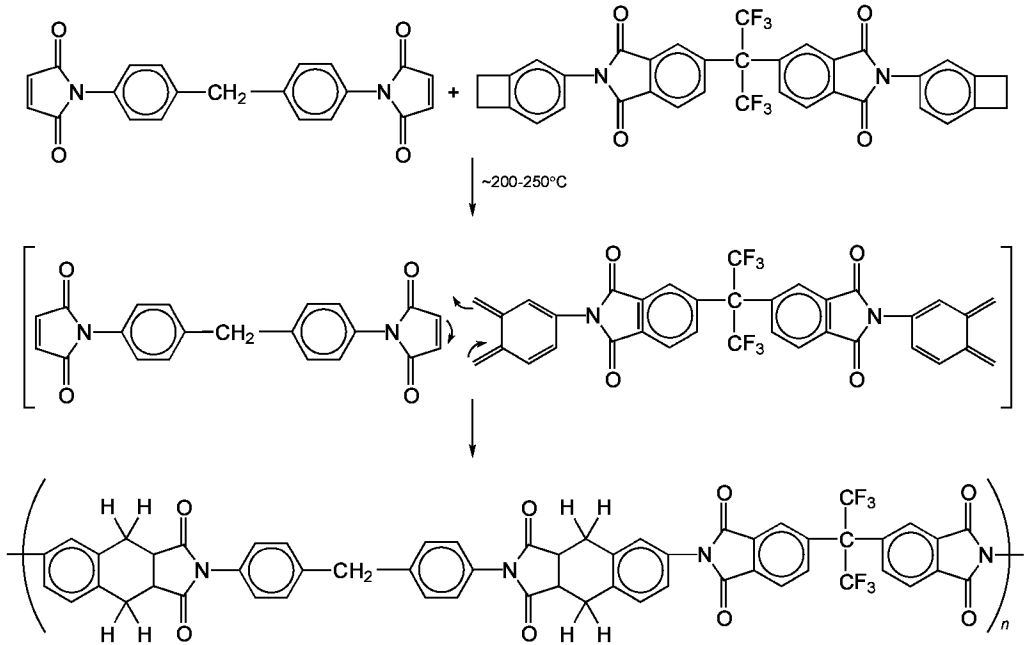


Fig. 10. Bismaleimide–benzocyclobutene copolymerization.

The resin system V-378A, mentioned earlier, is a bismaleimide system that has been modified with divinylbenzene to achieve drape and tack in prepreg form. Divinylbenzene-modified BMI is appreciated because of its outstanding hot–wet environmental resistance and epoxylike cure (18).

BMI–divinylbenzene systems cure via a Diels–Alder reaction.

Modified Bismaleimides. Bismaleimide resins may be further modified and blended with other thermoset resins or reactive diluents to achieve either specific end-use properties or processibility. Thermoset resins that can be used for modification are unsaturated polyesters, vinylesters, cyanate esters, and epoxies.

BMI–Epoxy Modifications. Bismaleimide–diamine prepolymers poly(amino bismaleimides) as described earlier can be used as curing agents for epoxy resins (52). Polyfunctional epoxy–novolac resin (EPIKOTE Resin 154) and tetraglycidyl-tetraphenylethane (EPON Resin 1031 [7328-97-4]) have been cured with this BMI. Heat-resistant, solventless resins comprising an epoxy resin, anhydride curing agent, and BMI resin are claimed in a U.S. patent (53). COMPIMIDE 795, a BMI–*m*-aminobenzoic hydrazide adduct, has been blended with tetraglycidyl-methylenedianiline (TGMDA) and 3,3′-diaminodiphenyl sulfone [27133-91-1] for use in hot-melt carbon fiber prepreg (43).

Heat-resistant resin compositions based on bismaleimide–epoxy blends are achieved by reaction of a BMI–*m*-aminophenol [591-27-5] (1:1) adduct with epoxy. This prepolymer is fully cured with an imidazole catalyst (54). Blends of hydroxy-terminated BMI–aminophenol adducts can easily be B-staged, that is, prepolymerized, and subsequently ground to provide a powder that can be molded by the application of heat and pressure.

A modified BMI–epoxy resin system has been introduced by Shell Chemical Company. The system is a highly reactive blend of a bismaleimide, COMPIMIDE 1206 (55–60% by weight solution of BMI in DMF), and EPON Resin 1151, a polyfunctional epoxy resin (55). In contrast with many polyimide resins on the market, no free methylenedianiline (MDA) is present in the product. This is an important feature, since MDA has been identified as an animal carcinogen and possible human carcinogen. This resin system has been fully evaluated for use in multilayer PCB boards (56). 2-Methylimidazole [693-98-1] is recommended as a catalyst. However, if required, the processing window can be widened by using 2-phenylimidazole [670-96-2], which is a more latent catalyst. There is a wide window for lamination of this system, from hot start, single pressure to dual pressure, vacuum assistance. The formulation consisting of 70 parts of COMPIMIDE 1206-R60 and 30 parts of EPON Resin 1151-BH 60 has an excellent property profile for the manufacture of multilayer circuit boards. Typical properties of COMPIMIDE 1206–EPON 1151 electrical laminates are compiled in Table 11.

Table 11. Properties of BMI–Epoxy^a Electrical Laminates

Properties	Value	
laminate composition, BMI–epoxy	70 30	50 50
flexural strength, 23°C, MPa ^b	462	441
flexural modulus, 23°C, GPa ^c	22	21
dielectric constant, 23°C	4.32	4.48
dissipation factor, 23°C	0.0083	0.0098
dielectric strength, V μm	30	31
volume resistivity, Ω·cm	2.1 × 10 ¹⁶	1.9 × 10 ¹⁶
surface resistivity, Ω·cm	3.8 × 10 ¹⁶	>1.9 × 10 ¹⁶
flammability, UL-94	V–O	V–O
copper peel, J m ^{2d}	1208	1348
methylene chloride, mg uptake	1.1	1.7
water absorption, wt %		
103 kPa ^e steam, 1 h.	0.28	0.31
boiling, 24 h	1.32	1.24
solder shock ^f	pass	pass
etchability ^g	0.03	0.04
tga, 5 wt % loss in air, °C	360	335

^a COMPIMIDE 1206–EPON 1151.

^b To convert MPa to psi, multiply by 145.

^c To convert GPa to psi, multiply by 145,000.

^d To convert J m² to lbf in., divide by 175.

^e 101.3 kPa = 1 atm.

^f 20s at 268°C 1 h after 103 kPa^e steam.

^g 5 min in conc H₂SO₄.

BMI–Biscyanate Modification. Based on the triazine technology developed by Bayer AG in Germany in the late 1960s (57), Mitsubishi Gas Oil Chemicals created a new family of resins that are blends of bismaleimides with biscyanates or biscyanate prepolymers. Biscyanate resins are also known as triazine resins, triazine referring to the chemical structure of the polymerized cyanate. Blends of bismaleimides and triazines are commercialized as BT resins (B-bismaleimide, T-triazine). There is no scientific literature available that deals with the chemistry of their copolymerization. The product manual for the commercial products claims that the copolymerization results in structures such as those given in Figure 11.

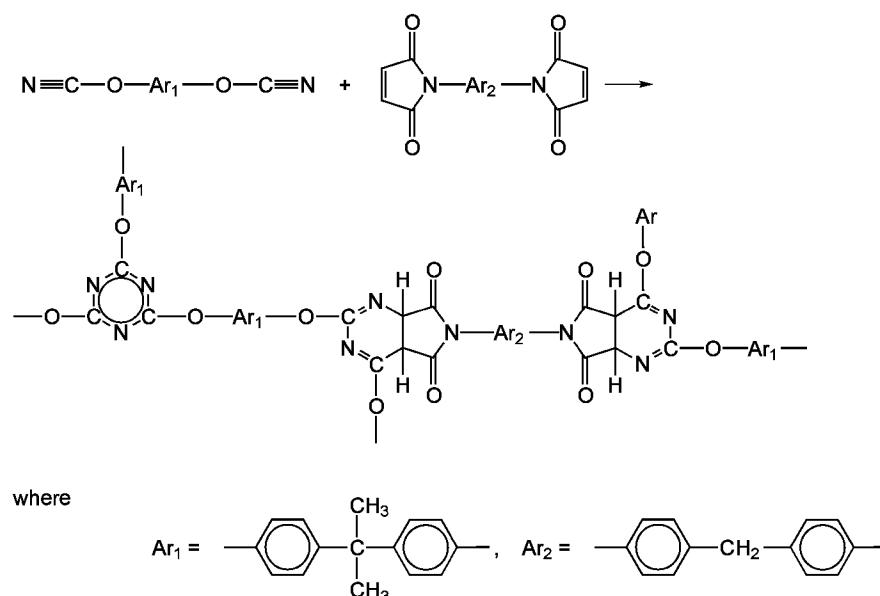


Fig. 11. Proposed copolymerization mechanism for bismaleimides with dicyanates.

A great variety of resin formulations is possible because other thermosets, such as epoxies or acrylates, and reactive diluents, such as *o*-diallyl phthalate [131-17-9], triallyl cyanurate [101-37-1], or triallyl isocyanurate [1025-15-6], can be used to further modify the BT resins. The concept is very flexible because bismaleimide and biscyanate can be blended and copolymerized in almost every ratio. If bismaleimide is used as a major constituent, then homopolymerization of the excess bismaleimide takes place in addition to the copolymerization. Catalysts such as zinc octoate or tertiary amines are recommended for cure. BT resins are mainly used in printed circuit and multilayer boards (58).

Toughened Bismaleimide Resins. Bismaleimide homopolymers are brittle thermosets and laminates made from them display low impact damage tolerance. The techniques used to toughen brittle bismaleimides include: the use of monomers or monomer–comonomer systems that provide inherently tough networks through reduced cross-link density; the use of elastomeric materials as modifiers to achieve second-phase toughening similar to rubber-modified polystyrene; and the use of engineering thermoplastics instead of elastomers to maintain the high elastic modulus of the thermoset in the toughened system.

BMI comonomer (allylphenols, propenylphenoxy compound) systems are significantly tougher than BMI homopolymers or BM-Michael-addition copolymer systems. BMI may also be toughened through elastomers and thermoplastics.

Rubber-Toughened Bismaleimides. Toughening of BMIs with carboxy-terminated acrylonitrile–butadiene rubbers (Hycar 1300 × 8 from BF Goodrich [68891-46-3]) has been reported (59). In contrast to epoxy resins, Hycar 1300 × 8 is not compatible with COMPIMIDE 353 bismaleimide at room temperature, and even at an elevated temperature (110°C) the rubber is not soluble in the BMI resin. However, during cure, the rubber chemically reacts with the BMI resin (via the double bonds present in the CTBN-rubber backbone) and becomes compatible to the extent that a microphase-separated morphology similar to the CTBN-rubber modified epoxy is achieved. The influence of the rubber concentration on the fracture energy G_{Ic} is presented in Figure 12. A significant improvement is achieved when rubber concentrations in excess of 50 phr are used.

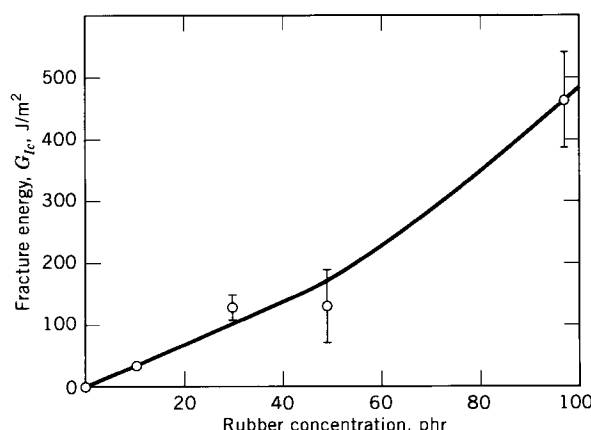


Fig. 12. Influence of rubber concentration on fracture energy (G_{Ic}). To convert J/m^2 to $ft\ lbf/in.^2$, divide by 2100.

The properties of CBTN-rubber modified BMI depend on the chemical composition of the rubber (60). The acrylonitrile content influences BMI–CTBN-rubber compatibility and therefore the morphology and the fracture energy of the modified resin. High acrylonitrile rubbers are more soluble and compatible with the BMI resin and provide a lower degree of phase separation and thus lower toughness. The principal drawback of CBTN-rubber toughened BMIs is the significant loss of high temperature properties, such as glass-transition temperature (T_g) and elastic modulus. As expected, the long-term aging stability at 200°C is poor.

Thermoplastic Toughening of Bismaleimides. Thermoplastics are increasingly used for the toughening of bismaleimide resins to improve their damage tolerance without compromising other important properties, such as glass transition temperature (T_g) and modulus of elasticity. High T_g thermoplastics such as polyhydantoins (61,62) (Resistofol N, Resistherm PH-10, Bayer AG), polyether-sulfones (51) (P1700, UCC), polyimides (53) (Matrimide 5218, CIBA GEIGY), polyetherimides (Ultem 1000, General Electric), and a series of poly(arylene ethers) (63), have been used to toughen BMI resins. All these thermoplastics, except the polyhydantoin PH-10, provided polyphase morphologies, and, at approximately 20% by weight of thermoplastic, phase inverted systems with the thermoplastic as the continuous phase are obtained. The influence of thermoplastic content on the fracture toughness of the cured BMI–thermoplastic blend system has been investigated, and, as expected, higher thermoplastic levels provide increased toughness. However, the various thermoplastics show different degrees of toughening (Table 12).

Table 12. Properties of Thermoplastic-Modified BMI^a Resins

Property	Control ^a	20 wt ^o _o PH-10 ^b	20 wt ^o _o RN ^c	20 wt ^o _o M 5218 ^d	20 wt ^o _o P 1700 ^e	13 wt ^o _o Ultem ^f	20 wt ^o _o PAE-1 ^g	20 wt ^o _o PAE-3 ^g	20 wt ^o _o PAE-4 ^g
<i>T_g</i> , °C	316	294	280	284	200	220	159.7	180.3	213.4
dry flexural properties									
room temperature									
strength, MPa ^h	115	115	118		0.5	117	114	124	132
modulus, GPa ^h	3.86	3.65	3.51	3.75	3.49	3.72	3.65	3.65	3.84
elongation, % ^o	3.05	3.09	3.52		2.67	3.35	3.21	3.40	3.64
120°C									
strength, MPa ^h		95	94		37	97	84	108	95
modulus, GPa ^h		2.88	2.89	3.13	1.93	3.02	2.60	3.04	3.00
elongation, % ^o		3.44	3.52		2.70	3.38	4.06	4.50	3.28
177°C									
strength, MPa ^h	84	91	78			45	26	30	100
modulus, GPa ^h	3.27	2.77	2.38	2.64		1.71	0.64	0.66	3.08
elongation, % ^o	2.73	4.52	4.74			2.99	6.61	>7.20	4.27
fracture toughness, <i>K_{IC}</i> , kPa√m	573	822	1322	929	763	805	1521	1235	1025
fracture energy, <i>G_{IC}</i> , J m ⁻²	85	185	498	230	167	174	634	418	274
moisture gain, ¹ % ^o	2.72	3.01		3.23			2.88	2.77	3.05

^a COMPIMIDE 796–COMPIMIDE 123.

^b Resisttherm PH-10 polyhydantoin.

^c Resistofol N polyhydantoin.

^d Matrimide 5218 polyimide.

^e Polyethersulfone.

^f Polyetherimide.

^g Poly(arylene ether)s.

^h To convert MPa to psi, multiply by 145.

¹ To convert J m² to filbf m², divide by 2100.

¹ Moisture gain was obtained on test specimens after 500 h at 94° o rh.

Many factors contribute to the toughness of a polyphase BMI–thermoplastic system, such as solubility parameters, phase adhesion, phase morphology, particle size, and particle size distribution. Another important factor is the molecular weight of the thermoplastic modifier. It has been demonstrated for a particular poly(arylene ether) backbone that high molecular weights increase the toughness of the blend system more than the low molecular weight counterparts (64). The toughness of the neat thermoplastic is likely to be a key factor in the corresponding thermoplastic–BMI blend.

Thermoplastic-modified BMIs are mainly used as matrix resins for advanced composites; therefore it is important that the good BMI-TP toughness in neat resin form translates into the composite and thus improves damage tolerance as measured via compressive strength after impact (CAI). For a BMI–polyetherimide system the toughness and morphology spectrum have been investigated, showing that for certain thermoplastic concentrations (>35% TP) the resin morphology is the same for the neat resin and the resin in the presence of the reinforcement. The thermoplastic is the continuous phase with very small (>0.5 nm) BMI inclusions (65). The fibers in the composite do not influence the development of the BMI–PEI morphology and, therefore, neat resin toughness correlates with composite toughness.

The principal disadvantage of the thermoplastic approach with high molecular weight polymers is the loss of processibility and the solvent sensitivity of thermoplastic–BMI blends. Therefore, attempts have been made to use reactive thermoplastics (functionalized thermoplastics) that can react into the BMI network and provide solvent resistance without loss of toughness. For a series of functionalized poly(arylene ethers), it has been demonstrated that a *M_w* (weight average molecular weight) of approximately 30,000 is required to maintain the toughness provided by the high molecular weight polyarylene ether. However, solvents do not completely destroy the blend system (64). Similar results have been obtained for BMI–diallylbisphenol A resin modified with amine-terminated poly(arylene ether sulfone) (66).

Toughening of BMIs with thermoplastics is a promising approach; however, more information is required about the toughening mechanism involved in order to select the most promising polymers in terms of backbone chemistry, molecular weight, and reactive groups.

Commercial BMI Resins. Since the late 1960s, bismaleimide resins have been commercially available. The first company in the marketplace was Rhône Poulenc Chimie, France. Several others, Technochemie, Germany; CIBA GEIGY, Switzerland; DSM, The Netherlands, BASF, West Germany; Mitsubishi Gas Chemicals and Mitsui Toatsu, both of Japan; Reichhold and Polyorganics from the United States; and Shell Chemical Company, United States (through the acquisition of Technochemie by Deutsche Shell AG), followed. Some of them lost interest in this speciality market sector. Table 13 indicates the BMI resins available at the end of 1992. No reference is made to prepreg systems available from various sources. A complete list of available products and information on any particular material should be obtained from the suppliers.

Table 13. Commercially Available BMI Resins

Company	Resin name	Description
CIBA GEIGY Corp.	Matrimide 5292 A,B	two-component resin system comprising 4,4'-bismaleimido-diphenylmethane (M5292A) and diallylbisphenol-A (M5292B)
	RD85–101	BMI building block based on diaminodiphenylindane; designed for hot-melt prepregging
	Araldite XU5292	bismaleimide resin solution for PCB applications
Du Pont Co.	MVA-2	1,3-bismaleimidabenzene building block
Mitsubishi Gas Chemicals Co.	BT resins	blends of bismaleimide(B) and triazine(T) resins; used primarily in printed circuit boards
Mitsui Toatsu Chemicals, Inc.	Bismaleimide-S	4,4'-bismaleimido-diphenylmethane. (The company also

Rhône Poulenc Chimie	Kerimide 601	offers other BMI building blocks.)
	FE70000	soluble BMI powder for multilayer board application
Shell Chemical Co. ^a	Rhodimid M3	series of research and development products
	COMPIMIDE MDAB	4,4'-bismaleimido-diphenylmethane
	COMPIMIDES 353, 353A, 796	high purity 4,4'-bismaleimido-diphenylmethane
	COMPIMIDE 15 MRK	unformulated BMI resins
	COMPIMIDE 65 FWR	for injection and compression molding
	COMPIMIDE 121	for filament winding
	COMPIMIDE 123	liquid bis(allylphenyl) compound
	COMPIMIDE 1206-F55	[bis-(<i>o</i> -propenyl)phenoxy] benzophenone
		55–60 wt % solution of bismaleimide for prepregging

^a Also, Technochemie GmbH-Verfahrenstechnik, Dossenheim, Germany is a member of the Royal Dutch Shell Group of Companies.

BIBLIOGRAPHY

1. Society of Plastics Industry, Committee on Resin Statistics, *Unsaturated Polyester Resins* 1988–1990.

2. *Modern Plastics* **68**(1) pp. 104, 108 (1991).

3. *Chem. Eng. News* **69**, No. 25 (June 24, 1991).

4. C. E. Browning, *Polym. Eng. Sci.* **18**, 16 (1978).

5. A. Apicella, L. Nicolais, G. Astarita, and E. Droli, *Polymer* **20**(9), 1143 (1979).

6. E. L. McKague, J. D. Reynold, and J. E. Haskin, *J. Appl. Polym. Sci.* **22**, 1643 (1978).

7. K. I. Hawthorn and F. C. Henson, "Epoxy Resin Chemistry II," in R. S. Bauer, ed., *ACS Symposium Series No. 221*, American Chemical Society, Washington, D.C., 1983, p. 135.

8. R. S. Bauer, *31st Int. SAMPE Symp.*, 1226 (April 7–19, 1986).

9. R. S. Bauer, *18th Int. SAMPE Tech. Conf.*, 510 (Oct. 7–9, 1986).

10. F. J. McGarry and A. M. Willner, *Dept. Civil Eng. Rep. R68-8*, MIT, Cambridge, Mass., 1968; F. J. McGarry and J. N. Sultan, *24th A. Tech. Conf. SPI 1969, Section 11-b*; F. J. McGarry and A. M. Willner, *ACS Div. Org. Coat. Plast. Chem. Prep.* **28**, Paper 512, 1968.

11. R. Siebert, "Rubber-Modified Thermoset Resins," in C. K. Riew and J. K. Gillham, eds., *ACS Advances in Chemistry Series 208*, American Chemical Society, Washington, D.C., 1983, p. 179; W. D. Bascom and D. L. Hunston, "Rubber Toughened Plastic," *Adv. Chem. Ser. No. 222*, American Chemical Society, Washington, D.C., 1989.

12. M. S. Sefton and co-workers, *19th Int. SAMPE Tech. Conf.*, 700 (1987); H. G. Recker and co-workers, *SAMPE Quarterly* **21**(1), 46 (1989).

13. **U.S. Pat. 3,380,964** (1968), F. Grundschober and J. Sambeth.

14. **Brit. Pat. Specification 1,190,718** (1968), M. Bargain, A. Combet, and P. Grosjean.

15. **U.S. Pat. 2,818,405** (1957), P. Kovacic.

16. H. D. Stenzenberger, in D. Wilson, H. D. Stenzenberger, and P. M. Hergenrother, eds., *Polyimides*, Blackie & Sons, Ltd. 1990, pp. 79–128.

17. **U.S. Pat. 3,966,864** (1976), H. D. Stenzenberger.

18. S. Street, *25th Nat. SAMPE Symp.* **25**, 366 (1980); **U.S. Pat. 4,351,932** (1982), S. Street; **U.S. Pat. 4,454,283** (1984), S. Street.

19. L. McKague, *28th Nat. SAMPE Symp.* **38**, 640 (1983).

20. B. L. Riley, *2nd Int. Conf. on Fibre Reinforced Composites, Proceedings*, University of Liverpool, UK, 1986, p. 153.

21. **Swiss Pat. Appl. 7988** (1975). S. Zahir and A. Renner.

22. **U.S. Pat. 3,127,414** (1964), H. N. Cole and W. F. Gruber.

23. C. K. Sauers, *J. Org. Chem.* **34** (8), 2275 (1969).

24. B. Lee, M. Chaudhari, and T. Gavin, *17th Nat. SAMPE Tech. Conf.* **17**, 172 (1986).

25. **U.S. Pat. 4,276,344** (1974), G. T. Kwiatkowski and G. L. Brode.

26. G. D. Lyle and co-workers, *ACS Polym. Prep.* **28**(2), 77 (1987).

27. **Ger. Offen. 2,031,573** (1971), F. F. Holub and M. L. Evans.

28. H. D. Stenzenberger in Ref. 16, pp. 85–87.

29. H. D. Stenzenberger, *J. Appl. Polym. Symp.* **22**, 77 (1973).

30. L. McKague, in *Composites for Extreme Environments, ASTM STP768* (1982).

31. **U.S. Pat. 3,562,223** (1971), M. Bargain, A. Combet, and P. Grosjean.

32. *Bismaleimides for Advanced Printed Circuit Boards*, N. N. Rhône Poulenc.

33. **U.S. Pat. 4,211,861** (1980), H. D. Stenzenberger.

34. H. D. Stenzenberger, W. Rümer, M. Herzog, M. S. Canning, and S. Pierce, *ICP, Technical Review* **28** (1) (Jan. 1987).

35. H. D. Stenzenberger, and co-workers, *32nd Int. SAMPE* **32**, 44 (1987).

36. **Jap. Kokai, 78 01 297** (1978), M. Kiyoji and S. Akira.

37. **Ger. Offen., 2,728,843** (1977), M. Kiyoji and O. Tsutomu.

38. P. M. Hergenrother, S. J. Havens, and J. Connell, *ACS Polym. Prep.* **27**(1) 409 (1986).

39. J. E. White, M. D. Scaia, and D. A. Snider-Tung, *ACS Polym. Prep.* **26**(1), 132 (1985).

40. J. W. Connell, R. G. Bass, and P. M. Hergenrother, *33rd Int. SAMPE Symp.* **33**, 251 (1988).

41. J. King, M. Chaudhari, and S. Zahir, *29th Nat. SAMPE Symp.* **29**, 392 (1984).

42. H. D. Stenzenberger and P. Künig, *High Performance Polymers* **1**(3), 239 (1989).

43. H. D. Stenzenberger, P. Künig, *High Performance Polymers* **1**(2), 133 (1989).

44. H. D. Stenzenberger, P. Künig, W. Rümer, M. Herzog, and W. Breitigam, *36th Int. SAMPE Symp.* **36**, 1236 (1991).

45. H. D. Stenzenberger and co-workers, *18th Int. SAMPE Tech. Conf.* **18**, 500 (1986).

46. H. D. Stenzenberger and co-workers, *32nd Int. SAMPE Symp.* **32**, 44 (1987).

47. **U.S. Pat. 4,789,704** (1988), H. D. Stenzenberger and P. Künig.

48. **U.S. Pat. 4,902,778** (1990), J. Boyd.

49. H. D. Stenzenberger and P. Künig, *High Performance Polym.* **3**(1) (1991).

50. Loon-Seng Tan, F. E. Arnold, and E. J. Solosky, *J. Polym. Sci., Polym. Chem. Ed.* **26**, 3103 (1988).

51. Loon-Seng Tan, F. E. Arnold, and E. J. Solosky, *Polym. Mat. Sci. Eng.* 650 **56**, (1987).

52. **Ger. Offen., 2,026,423** (1970), M. Bargain and M. Gruffaz.

53. **U.S. Pat. 3,730,948** (1973), K. Akiyama and K. Makino.

54. **Jpn. Kokai 780,1226** (1978), M. Kiyoji and S. Akino.

55. A. M. Pigneri, E. C. Galgoci, R. J. Jackson, and G. E. Young, *1st Int. SAMPE Electr. Conf.* **1**, 657 (1987).

56. M. J. Davis and T. R. Sense, *ICP Fall Meeting*, Chicago, Ill., 1987.

57. R. Kubens, H. Schultheis, R. Wolf, and E. Grigat, *Kunststoffe* **58**, 827 (1968).
58. S. Ayano, *Kunststoffe* **75**, 475 (1985).
59. S. J. Shaw and A. J. Kinloch, *Int. J. Adhesion Adhesives* Vol **5**(3), 123 (1985).
60. Sh. Takeda and H. Kakiuchi, *J. Appl. Polym. Sci.* **35**, 1351–1366 (1968).
61. H. D. Stenzenberger, W. Rümer, M. Herzog, and P. Kunig, *33rd Int. SAMPE Symp.* **33**, 1546 (1988).
62. H. D. Stenzenberger, W. Rümer, M. Herzog, and P. Kunig, *34th Int. SAMPE Symp.*, closed session.
63. H. D. Stenzenberger, W. Rümer, P. M. Hergenrother, and B. Jensen, *34th Int. SAMPE Symp.* **34**, 2054 (1989).
64. H. D. Stenzenberger, W. Rümer, P. M. Hergenrother, B. Jensen, and W. Breitigam *35th Int. SAMPE Symp.* **35**, 2175 (1990).
65. D. Rakutt, E. Fitzer, and H. D. Stenzenberger, *High Perform. Polym.* **3**(2), 59 (1991).
66. S. P. Wilkinson, S. C. Liptak, P. A. Wood, J. E. McGrath, and T. C. Ward, *36th Int. SAMPE Symp.* **36**, 482 (1991).

Ronald S. Bauer
Steven L. Stewart
Shell Development Company

Horst D. Stenzenberger
Technochemie GmbH-Verfahrenstechnik

THERMOPLASTICS

Thermoplastic resins have received considerable attention in the past few years as the matrix material in organic resin-based composites. Whereas thermosetting materials date back 5000 years, when Egyptians used a straw-reinforcing agent in a clay matrix to form bricks, thermoplastic composites are relatively recent. Although their use in advanced composites is not widespread, thermoplastic composites are used extensively in commercial applications ranging from automobiles to durable goods (1) (see ENGINEERING PLASTICS).

Thermoplastic polymers are usually linear molecules with no chemical linkage between the molecules. The molecules are held together by weak secondary forces, such as van der Waals or hydrogen bonding, and as such are readily deformed by the application of heat or pressure. Thermoplastic resins can be amorphous, that is, structureless, or semicrystalline, in which some of the molecules form an ordered array. A material is usually considered semicrystalline if as little as 5% of the polymer is in the crystalline form. Semicrystalline resins exhibit a higher modulus, but amorphous materials are tougher; amorphous materials are usually more solvent sensitive but can be processed at lower temperatures. One of the most important advantages of thermoplastic resins is their toughness, that is, high impact strength and fracture resistance, which, unfortunately, is not linearily translated into properties of the composite. Other advantages of thermoplastic polymers include long shelf life at room temperature; postformability, that is, thermal reforming; ease of repair by thermal welding or solvents; and ease of handling, that is, they are not tacky.

Although in a broad sense, a composite is a material that contains two or more distinct phases, this discussion is restricted to those materials in which one of the phases, usually the fibers, serves as a reinforcing or strengthening agent (see COMPOSITE MATERIALS, SURVEY). Materials in which a relatively inexpensive material, such as talc or calcium carbonate, is added to reduce the overall cost of the expensive thermoplastic polymer are not included. Advanced or high performance composites contain more than 50 wt % fibers, whereas other composites or commodity composites (reinforced plastics) contain 20–40 wt % reinforcing agent. The distinction is real, and different resin properties, hence different chemistries, are required for these materials. In general, the specific chemistry of the resin system cannot be separated from the processing method used to prepare the final composite. Although the fundamental structure of the polymer is the same regardless of the method used to synthesize it, significant differences in the form and reactivity of the final material can be observed.

Because the early thermoset resins used in advanced composites were quite brittle and tougher resins were needed, thermoplastic resins seemed to have great potential as composite matrix materials. In addition to being tough, they have a long shelf life and are thermally reformable. However, the primary user of advanced composites, the aerospace industry, has a large investment in capital equipment that utilizes tacky, drapable, carbon–epoxy composites, which are processed as prepregs at temperatures from 121° to 177°C at high pressures in an autoclave. Considerable effort has been expended to develop thermoplastic resins that can be processed with the existing equipment, that is, resins that can be (1) coated directly onto fibers to prepare conventional prepregs; (2) spun into thermoplastic resin fibers that can be woven together with the reinforcing fiber, that is commingled, and subsequently consolidated; (3) cured from a partially polymerized polymer and coated directly onto the fiber to form a prepreg similar to that used for thermosets; or (4) prepared as small particles that can be coated onto the reinforcing fiber from either solution or by electrostatic attraction and subsequently heated to form a continuous phase on the fiber.

One of the principal advantages of true thermoplastic polymers is their ability to consolidate or flow at elevated temperatures; however, this quality also limits their upper-use temperature. Amorphous materials begin to flow or creep above the glass-transition temperature, T_g , whereas crystalline resins must be heated above the melting point, T_m . As a rule of thumb, T_g is approximately $2/3 T_m$ (Kelvin temperature), so that crystalline polymers may begin to degrade at the temperatures required for processing. Even a relatively low temperature material such as polypropylene, which melts between 168 and 175°C must be heated to approximately 180–190°C to process the composite. In many instances the processing window is quite narrow owing to a lower temperature limit set by the melting point and an upper temperature limit set by the rate of thermal degradation.

Much of the recent work on the synthesis of new resin systems has centered on two important problems: (1) methods to form thermoplastic resins at relatively low temperatures while maintaining high temperature capability; and (2) development of low viscosity resins that readily flow into and around an existing preform structure but rapidly polymerize to a stable thermoplastic resin when heated. For example, materials with reactive end groups or cyclic compounds that readily form high molecular-weight polymers during processing are being developed for both conventional and high performance composites.

In spite of the many advantages of and large data base on thermoplastic resins, the development of advanced thermoplastic composites has been much slower than the development of thermoset composites. Because of their high melt or solution viscosities, the incorporation of the reinforcing phase into the thermoplastic matrix has limited the development of thermoplastic composites. A great deal of work has been devoted to the development of materials and processing methods to circumvent these problems. Although fibers can be coated in several ways, it is necessary to prepare the polymer in a form suitable to the particular processing method. Thus it is necessary to prepare a given generic material by a different synthesis route in order to obtain the final resin in the desired form.

Thermoplastic composites can be classified according to use, cost, performance, or processing methods. In the following discussion of the chemistry of the resin systems utilized in composites, three classes are considered:

- (1) Resins used in conventional composites (usually commodity materials) and typically containing 10–40 wt % reinforcing agent.

(2) Resins used in advanced or high performance composites, which contain more than 50 wt % reinforcing agent and are the typical aerospace materials.

(3) Pseudothermoplastic resin systems, which are formed as conventional thermoplastic materials and then cured or postcured in a manner similar to that used for thermosetting resins to enhance high temperature properties.

Commodity Resins

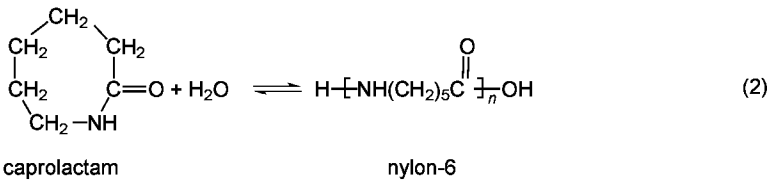
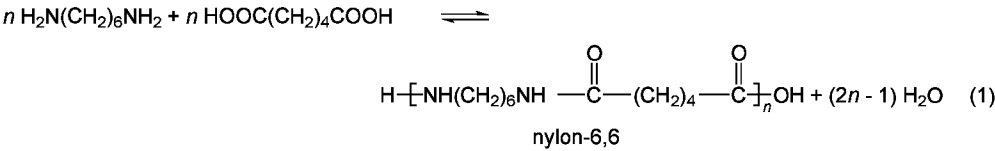
Most of the resin systems used in commodity composites are slight modifications of the standard commercial molding grade material. Usually certain selected properties, such as purity or molecular weight range or distribution, are enhanced or carefully selected. In addition, special additives, such as flow controllers, thermal stabilizers, or antioxidants, are often added by the resin manufacturer prior to shipment. Many of the conventional or commodity-type resins used in thermoplastic composites are listed in Table 1 and the preparation of each of these is described. All resins and blends described in the literature are not listed, and the synthesis described is not the only procedure available, but is usually the most common commercial process.

Table 1. Conventional Thermoplastic Resins

Chemical class	Repeat unit	Polymer	Morphology
polyamides		nylon-6,nylon-6	semicrystallinesemicrys

			talline
polyolefins		polypropylenepolystyrene and its copolymers	semicrystallineamorphous
polycarbonates		polycarbonate	amorphous
polyesters		poly(ethylene terephthalate)poly(butylene terephthalate)	semicrystallinesemicrystalline
acetals		polyoxymethylene	semicrystalline

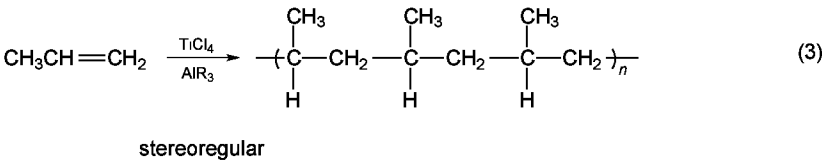
Polyamides. Nylon was the first commercial thermoplastic polymer produced on a large scale, as a direct result of the pioneering work of Carothers (2). Nylons contain repeating amide functionality (see POLYAMIDES). The aliphatic polyamide nylon is one of the most important thermoplastic materials (3). It is used extensively as a molding compound and in fiber glass (both short and long fibers) reinforced composites. Nylon-6,6 [32131-17-2] and nylon-6 [25038-54-4] were first put into commercial production in 1938 and 1939, respectively. In view of the 50 years of commercial experience with nylons, a great deal is known about the chemistry of these materials and the various additives used to enhance specific properties. However, much of the specific knowledge remains proprietary information of the various suppliers. The molding grades form the basis for the composite grade material. There are two principal synthesis routes for the production of nylon: condensation of a diamine and dibasic acid (eq. 1) or rearrangement of a lactam (eq. 2).



One of the primary factors in the use of nylons to prepare composites is their melt viscosity, which is controlled by the proper choice of molecular weight. Typical commercial nylons are available with molecular weights ranging from 11,000 to 40,000. The molecular weight chosen for a particular composite application represents a compromise between a low value, yielding a low viscosity to enhance fiber wetting, and a high value, to increase toughness and promote maximum crystallinity.

Many modifiers and additives have been described for use with nylon composites, but generally a small amount, 0.05–1 wt %, of a lubricity aid, such as sodium or zinc stearate (4) is added to enhance both resin flow during processing and removal from the mold after consolidation.

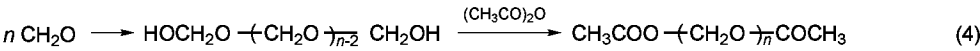
Polyolefins. The most common polyolefin used to prepare composites is polypropylene [9003-07-0], a commodity polymer that has been in commercial production for almost 40 years following its controlled polymerization by Natta in 1954 (5). Natta used a Ziegler catalyst (6) consisting of titanium tetrachloride and an aluminum alkyl to produce isotactic polypropylene directly from propylene:



Polypropylene is available with many different reinforcing agents or fillers, such as talc, mica, or calcium carbonate; chopped or continuous strand fiber glass is the most common reinforcing agent used for composites. Many additives have been developed to enhance the thermal stability of polypropylene to minimize degradation during processing. One of the most important requirements of the polypropylene used in the manufacture of composites is that it be relatively pure and free of residual catalyst. Recent developments to form copolymers of polypropylene and polyethylene have great promise for relatively inexpensive, tough, thermoplastic composite applications (see OLEFIN POLYMERS).

Another polyolefin of interest is polystyrene, a clear, brittle plastic that, by itself, is rarely used in composites. However, several copolymers and alloys of polystyrene with acrylonitrile or butadiene have been used with fiber glass or glass spheres to form composites (7).

Acetals. Acetal resins (qv) are polymers of formaldehyde and are usually called polyoxymethylene [9002-81-7]. Acetal homopolymer was developed at Du Pont (8). The commercial development of acetal resins required a pure monomer. The monomer is rigorously purified to remove water, formic acid, metals, and methanol, which act as chain-transfer or reaction-terminating agents. The purified formaldehyde is polymerized to form the acetal homopolymer; the polymer end groups are stabilized by reaction with acetic anhydride to form acetate end groups (9).

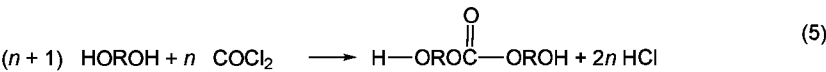


The polymer also can be made from trioxane (the trimer of formaldehyde), usually as a copolymer with ethylene oxide. The —CH₂CH₂— fragments in the copolymer chain prevent depolymerization; acetal copolymer was developed by Celanese (10).

Precise amounts of chain terminators are added during polymerization to produce final polymers having various molecular weights. Stabilizers and antioxidants are added to both the homopolymer and the copolymer to create the basic polymer grades. Although several fillers, such as talc or calcium carbonate, can be added to polyoxymethylene, the most common reinforcing agent for engineering applications is fiber glass. Reference 11 provides an

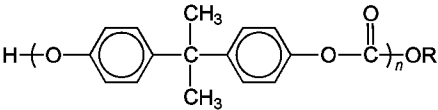
excellent review of various fillers and reinforcing agents used with acetals.

Polycarbonates. Polyarylates are aromatic polyesters commonly prepared from aromatic dicarboxylic acids and diphenols. One of the most important polyarylates is polycarbonate, a polyester of carbonic acid. Polycarbonate composite is extensively used in the automotive industry because the resin is a tough, corrosion-resistant material. Polycarbonates (qv) can be prepared from aliphatic or aromatic materials by two routes: reaction of a dihydroxy compound with phosgene accompanied by liberation of HCl (eq. 5):



or transesterification of a dihydroxy compound with dialkyl or diaryl carbonates.

The most common commercial polycarbonate [24936-68-3] is prepared from 2,2-bis (4-hydroxyphenyl)propane, that is, bisphenol A [80-05-7], and has the general structure:

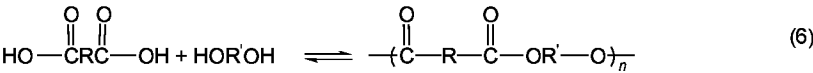


where R could be H or C H₅.

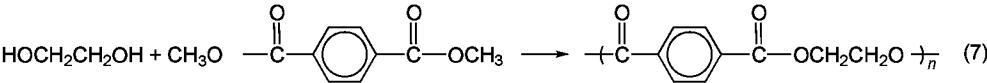
The principal industrial method of preparing this material is a two-phase reaction in which the polymerization reaction occurs at the interface (12,13). The disodium salt of bisphenol A in aqueous alkaline solution reacts with the phosgene in an inert chlorinated solvent, such as methylene chloride. The mono- or dichloroformates condense and in a second stage are polymerized to a high molecular-weight polymer using a basic catalyst, such as triethylamine. The degree of polymerization depends on the degree of mixing and the pH of the aqueous phase. The molecular weight of the polymer is controlled by adding a chain terminator such as phenol. The usual molecular-weight range used for the manufacture of composites is 20,000–35,000. When used to prepare composites, polycarbonate must be dry and free of both acid and basic impurities to minimize depolymerization and degradation during processing. Most polycarbonate composites contain between 5 and 30 wt % fiber glass reinforcing agent, but composites with up to 40 wt % can be made. E-glass is the material of choice because it is alkaline free.

Workers at the General Electric Co. have described a new procedure using low viscosity cyclic intermediates to prepare polycarbonate composites (14,15). This work is discussed in the section dealing with novel methods.

Polyesters. Polyesters (qv) are widely used as the matrix for conventional composites. Two resins of particular importance because of the large amounts used are (poly(ethylene terephthalate) [25038-59-9] (PET) and poly(butylene terephthalate) [24968-12-5] (PBT). Although polyesters can be made from diacids and diols by direct condensation,



they form relatively low molecular-weight polymers that have little commercial value. The most practical methods for the preparation of high molecular-weight polyesters involves ester-exchange reactions. PET is usually prepared by the reaction between ethylene glycol and the dimethyl ester of terephthalate acid in acid solution (16):



PET is a semicrystalline material with a nominal degree of crystallinity of approximately 60 vol %; however, because the rate of crystallization is relatively slow, it is desirable to add a nucleating agent, such as talc, to the resin (17). If a nucleating agent is not added, fabricated parts have a nonhomogeneous morphology and it is difficult to maintain dimensional stability. It is also possible to add a plasticizer, such as dibenzoate of neopentyl glycol, to lower the *T_g* and permit rapid crystallization throughout the part (18). Composites containing up to 40 vol % fiber glass are commercially available.

PBT is produced by the transesterification of dimethyl terephthalate with 1,4-butanediol by means of a catalyzed melt polycondensation (19). PBT is also semicrystalline and is an extremely tough resin. Several commercial resins use a blend of PBT with another resin, such as PET, polycarbonate, or nylon. Typically, composites of PBT contain 20–30 vol % fiber glass.

Advanced Composites

The term *advanced composite* normally implies a high volume fraction reinforcing agent, of the order of 60 wt %. Much research into the development of high temperature advanced thermoplastic composites has been conducted since the late 1970s. The primary goal of this work has been the development of high temperature tough materials. This development presents a twofold problem: (1) the synthesis of resins that are stable and do not degrade at the high temperatures of interest, and (2) incorporation of the resin system into a reinforcing agent so that a structural composite can be fabricated. The latter is the more difficult of the two problems.

Resins for advanced composites can be classified according to their chemistry; typical resins are polyaryletherketones, polysulfides, polysulfones, and a very broad class of polyimides containing one or more additional functional groups (Table 2) (see also ENGINEERING PLASTICS).

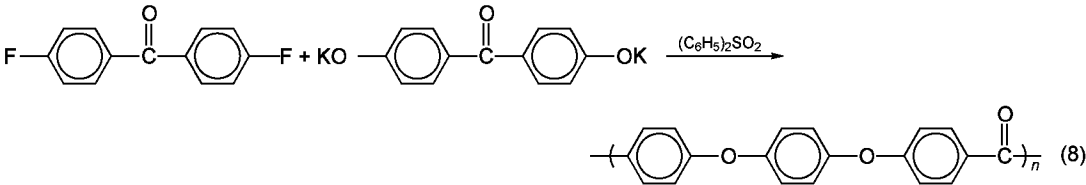
Table 2. Resins for Advanced Composites

Chemical class	Resin	Morphology	<i>T_g</i> , °C	<i>T_m</i> , °C
polyetherketones	polyether ether ketone	semicrystalline	143	345
	polyether ketone ketone	semicrystalline	170	370
polysulfones	polysulfone	amorphous	200	
	polyethersulfone	amorphous	230	
polysulfides	polyphenylene sulfide ^a	semicrystalline	85	285
polyimides	polyimide ^a	semicrystalline	250–300	
	polyetherimide	amorphous	220	
	polyamide imide ^a		275	
	polybenzimidazoles ^a		230–290	
	fluorinated polyimides	amorphous	340	

^a These materials are also referred to as pseudothermoplastics.

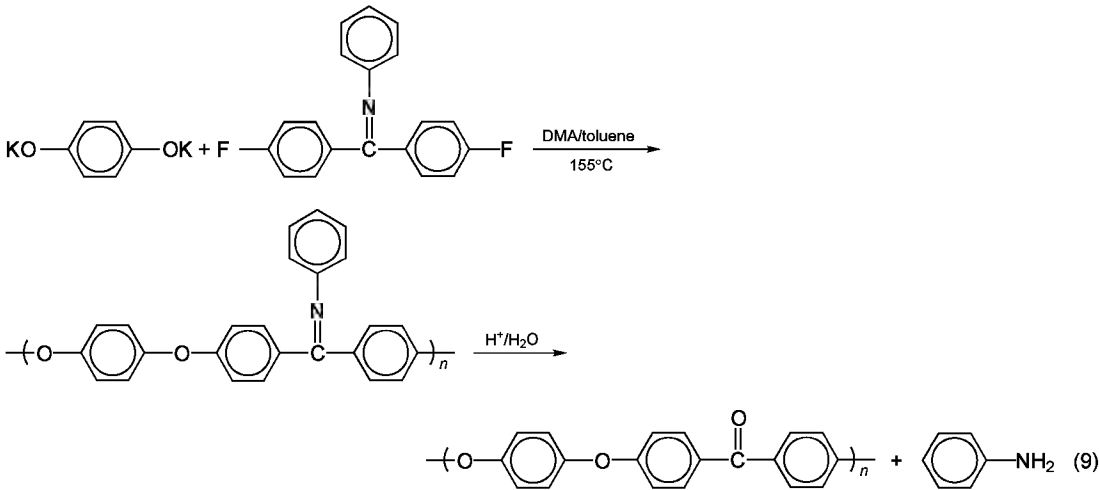
It is important to distinguish between true thermoplastics and pseudothermoplastics. Pseudothermoplastics are systems in which the chemistry continues during processing or an extra postcure step is added after the resin is consolidated. During the postcure process, the molecular weight may continue to increase, reactive pendent or end groups may cross-link, or volatile compounds such as reaction products or residual solvent may be expelled. In general, the properties change and can be likened to cross-linking or chain extension. Many of the so-called advanced thermoplastic resins are actually pseudothermoplastics. Although pseudothermoplastics are not completely reprocessable as are the true thermoplastics, usually they can be thermoformed a second or possibly a third time if necessary to correct defects in a structural component.

Polyarylether Ketones. The aromatic polyether ketones are true thermoplastics. Although several are commercially available, two resins in particular, poly ether ether ketone [31694-16-3] (PEEK) from ICI and poly ether ketone ketone (PEKK) from Du Pont, have received most of the attention. PEEK was first synthesized in 1981 (20) and has been well studied; it is the subject of numerous papers because of its potential use in high performance aircraft. Tough, semicrystalline PEEK is prepared by the condensation of bis(4-fluorophenyl) ketone with the potassium salt of bis(4-hydroxyphenyl) ketone in a diaryl sulfone solvent, such as diphenyl sulfone. The choice of solvent is critical; other solvents, such as liquid HF, promote the reaction but lead to premature low molecular-weight crystals, which do not exhibit sufficient toughness (21).



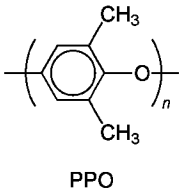
A commercial prepreg of PEEK and carbon fibers (manufactured and sold by ICI as APC-2) can be made by the hot melt process. Although the prepreg prepared in this manner has been used extensively, it is quite stiff and boardy and makes it difficult to form structures with complex shapes. The PEEK used to make the APC-2 composite is a high purity grade that contains no special additives. An excellent discussion of the effects of heat transfer during processing of APC-2 on the morphology of PEEK has been given (22).

More recently a method has been described to prepare fine (0.5–5 μm) particles of PEEK for use in powder prepregging (23). PEEK resin is too tough to be ground to a fine powder suitable for powder prepregging. This represents an excellent example where the choice of processing method, in this case the method of coating the resin onto the fiber, controls the synthesis procedure. In this synthesis (23) a fluorinated ketamine reacts with the potassium salt of dihydroxybenzene in dimethylacetamide–toluene solution at 155°C. The resulting ether ketamine polymer is hydrolyzed in acid solution to form PEEK polymer (eq. 9). The size and shape of the particles is controlled by the acid concentration (that is, rate of hydrolysis) and agitation during the hydrolysis step. Uniform crystalline particles as small as 0.5 μm can be obtained.

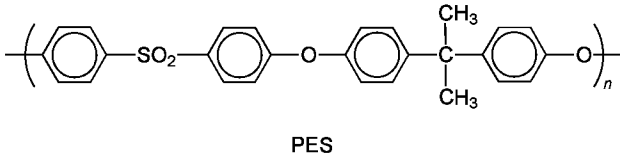


PEEK can also be spun into fibers, which are commingled with a reinforcing fiber to form a yarn (24). The PEEK fiber is not used as the reinforcing agent but, when heated above its melting point, flows around the reinforcing fibers (typically carbon) and forms the resin matrix. The commingled yarn is woven into the shape desired and consolidated at temperatures above the melting point of PEEK. The PEEK used in this material is resin grade, which is slightly less pure than the PEEK used in APC-2. The overall effect of this slight difference in purity between the resin used in APC-2 and the commingled yarn on the long-term performance of the composite is not known.

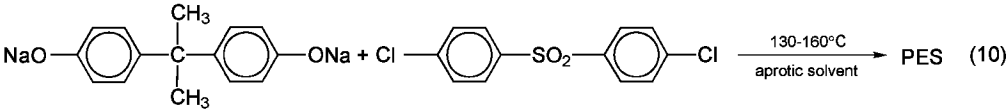
Poly(phenylene ether). The only commercially available thermoplastic poly(phenylene oxide) PPO is the polyether poly(2,6-dimethylphenol-1,4-phenylene ether) [24938-67-8]. PPO is prepared by the oxidative coupling of 2,6-dimethylphenol with a copper amine catalyst (25). Usually PPO is blended with other polymers such as polystyrene (see POLYETHERS, AROMATIC). However, thermoplastic composites containing randomly oriented glass fibers are available.



Polysulfones. The most common polysulfone is actually a sulfone ether, polyethersulfone [25135-51-7] (PES), and has the following structure.



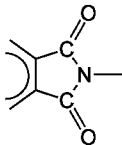
It is prepared from the polycondensation of the disodium salt of bisphenol A and 4,4-dichlorodiphenyl sulfone in a polar aprotic solvent such as dimethyl sulfoxide (26).



This material is a true amorphous thermoplastic with a T_g of 185°C and can be made into thermoforming composite with 20–40 wt % glass fiber (see POLYMERS CONTAINING SULFUR).

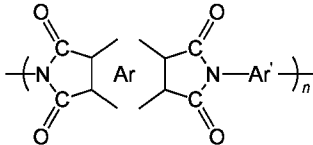
Polyimides. Polyimides (qv) represent a very broad class of materials ranging from thermosets to both amorphous and semicrystalline thermoplastics. Polyimides are used extensively as high temperature adhesives, as insulation for electrical wire, and for printed circuit boards. They have great potential for high temperature applications because of their thermo-oxidative stability, solvent resistance, and general ease of preparation. However, their use in resin matrix composites has been rather limited in spite of their great potential. Much has been written about polyimide resins for use in high temperature composites, but the amount of literature generated greatly exceeds their commercial value. Reference 27 is an excellent summary of some of the many studies on polyimides.

All polyimides contain the basic imide functionality:

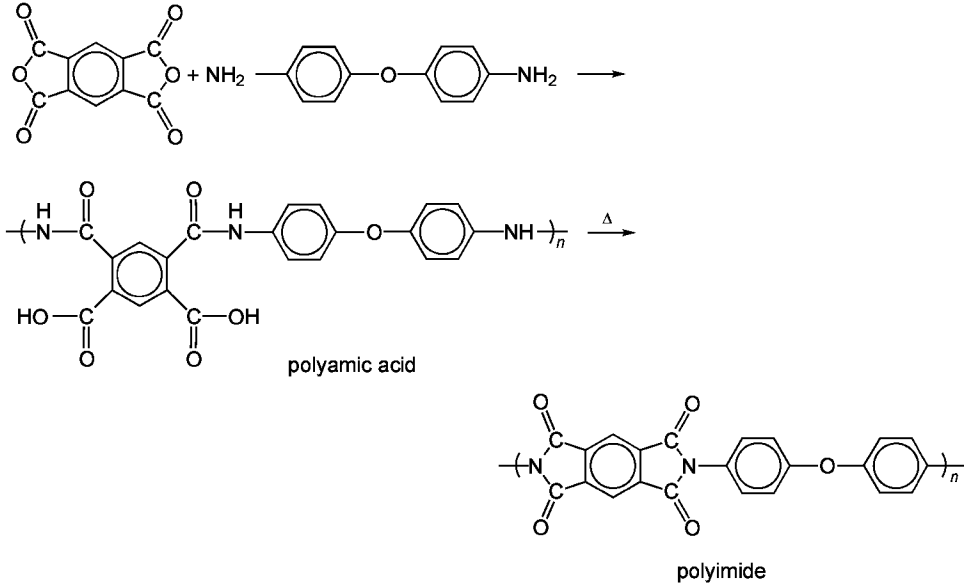


In addition they may contain ether, amide, carbonyl, sulfone, or other functional groups. References 28 and 29 provide excellent reviews of polyimide chemistry.

Aromatic polyimides are generally produced by the reaction of aromatic dianhydrides with aromatic diamines to yield a material with the general structure



where either the aromatic ring Ar or Ar' can contain a functional group. The first commercial polyimide, poly(4,4'-oxydiphenylpyromellitimide) [25036-53-7], manufactured by Du Pont and sold under the trade name Kapton, is a high temperature imide-ether polymer prepared from pyromellitic dianhydride and 4,4'-diaminodiphenyl ether according to the following reactions (30):

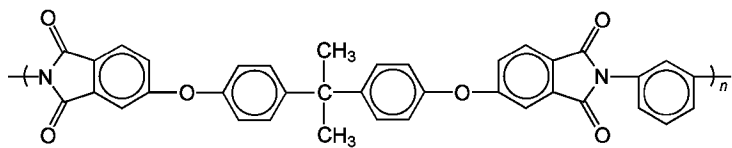


The polyimide shown is a true thermosetting resin, but the general reaction procedure, coupling the dianhydride with the diamine, is extremely important throughout polyimide chemistry. The intermediate polyamic acid polymers form the basis for many of the polyimide resins used in advanced composites.

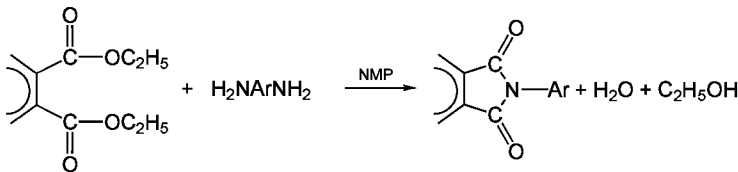
Different dianhydrides reacting with different diamines can be used to prepare a wide variety of polyamic acids. Although the polyamic acid is a high molecular-weight polymer, it is soluble in solvents such as *N*-methylpyrrolidinone (NMP) and in this form can wet the reinforcing agent prior to the final imidization step. Because of the high temperature capability of polyimides and the desirable overall properties of thermoplastic resins, considerable effort has been devoted to the development of thermoplastic polyimide-based materials (29).

Even though linear aromatic condensation types of thermoplastic polyimides can be prepared according to the reactions described and are, at least in principle, melt processible, they are quite difficult to process because of the extremely high (315–400°C) temperatures and high pressures required. In some instances the polyamic acid (in an appropriate solvent) is used and the final imidization step conducted during processing. In this case, both the solvent and the volatile compounds formed during ring closure must be eliminated from the consolidated composites. The total volatile content may amount to 15–20 wt % of the resin systems. To eliminate this undesirable reaction, considerable research has been devoted to synthesizing low molecular-weight imide oligomers with reactive end groups, such as maleimide, norbornene, or acetylene, which can be condensed during the processing stage. In many cases, these resins undergo a small amount of crosslinking during processing, and they are best referred to as pseudothermoplastics.

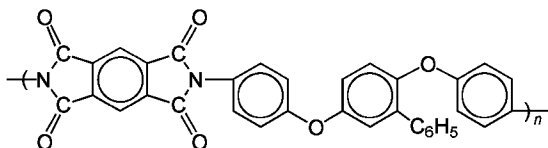
Polyether Imides. Polyether imides (PEIs) are amorphous, high performance thermoplastic polymers that have been in use since 1982. The first commercial polyether imides were the Ultem series developed by the General Electric Co. The first, Ultem 1000 [61128-24-3], is prepared from phthalic anhydride, bisphenol A, and meta-phenylenediamine and has the following structure:



PEI resins reinforced with up to 40 wt % fiber glass are available. The American Cyanamid Co. uses an Ultem-type resin with carbon fibers to form their Cypac 1000 series prepregs (31). The moisture content of the resin must be less than 0.05% prior to melt processing to minimize thermal degradation. Another series of PEI resins has been introduced by Du Pont, the Avimid series. The Avimid K-III [94148-82-0] resin was formulated to be a high temperature thermoplastic polyimide that could be processed with existing autoclave equipment designed for thermosets (32). The tack and drape of the prepreg are similar to those of the epoxies but the final resin is a tough thermoplastic. The exact structure and formulation has not been published but, in general, an aromatic diethyl ester of a diacid reacts with an aromatic diamine in a solvent such as *N*-methylpyrrolidinone (NMP), that is,

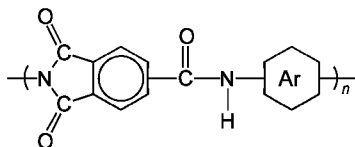


The structure of Avimid K-III is similar to



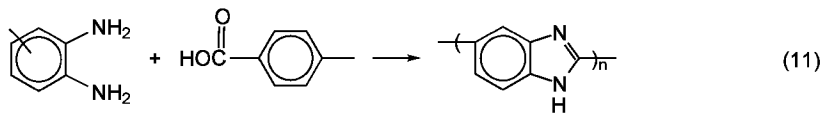
A significant problem associated with this resin system is the large amount, 14–16%, by weight of NMP solvent and condensation reaction by-products (water and ethanol). Special autoclave processing cycles are required to remove these volatile compounds from a finished part. During early stages of processing a semicrystalline polymer forms. The crystalline intermediate melts and the molten polymer consolidates into a tough amorphous material that does not crystallize upon cooling. The material is processed at approximately 360°C using a 1.5 MPa (200 psi) consolidation pressure. The final product is not a true thermoplastic but a pseudothermoplastic, which, however, can be postformed into simple geometries. Composites containing up to 60 wt % with either carbon or aramid reinforcing agents are available.

Polyamide Imides. Polyamide imides (PAIs) are formed from the condensation of trimellitic anhydride and aromatic diamines (33). The polymer is called amide–imide because the polymer chain comprises amide linkages alternating with imide linkages, with the general chemical structure:

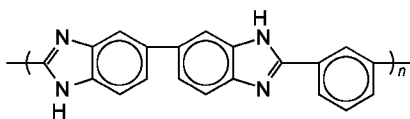


One commercial PAI is Tordon [42955-03-3], supplied by Amoco (33). Some PAI grades must be postcured after molding to fully develop their mechanical properties for long-term service and to keep parts from blistering or distorting during use at high temperatures. Several of the commercial grades of PAIs contain approximately 1 wt % polytetrafluoroethylene and can be used with either glass or carbon reinforcing fibers (34). The required postcure at 260°C can range from 5 to 30 days. Postcuring causes significant changes, most notably a large increase in molecular weight to form a pseudothermoplastic material, as noted previously.

Polybenzimidazoles. The polybenzimidazoles (PBIs) are generally produced by the high temperature, melt polycondensation reaction of aromatic bis-ortho-diamines and aromatic dicarboxylates (acids, esters, or amides) in a reaction such as that shown in equation 11 to form benzimidazole [51-17-2] as the repeating unit.



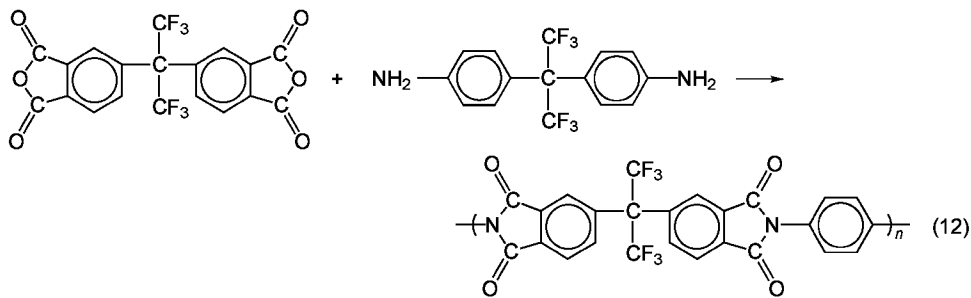
A large number of PBIs have been investigated since their first synthesis in 1961 (35); the particular polymer used in many commercial and developmental applications is poly(2,2'-*m*-phenylene)-5,5'-dibenzimidazole).



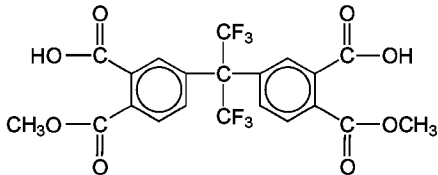
The distinction between thermoplastic and thermoset is particularly cloudy in the case of PBIs because they are formed and processed as thermoplastics but after processing exhibit properties that are more closely related to thermosetting materials. Hence they, too, are usually considered pseudothermoplastics.

Fluorinated Polyimides.

Often the substitution of fluorine atoms for hydrogen atoms in a polymer chain markedly increases the thermal stability of the base polymer; this is true for polyimides. A typical fluorinated polyimide is prepared from the reaction of 2,2-bis(3,4-dicarboxyphenyl)hexafluoropropane dianhydride and 2,2-bis-(4-amino phenyl)hexafluoropropane according to the following reaction (36):

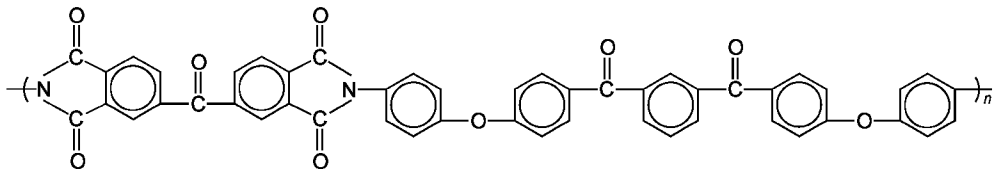


This material is manufactured by Du Pont and sold as resin NR-150B2. A carbon fiber prepreg using NR-150B2 is available and is a member of the Avimid series of composites, Avimid [70850-07-6]. The resin used in the prepreg also contains the following diester-diacid:



and a mixture of meta- and para-phenylenediamines with 5–10% solvent of NMP.

Semicrystalline Polyimides. Semicrystalline polyimides containing carbonyl and ether groups have been synthesized by the group at the NASA Langley Research Center (37). One such material, designated LARC-CPI, which is an acronym for Langley Research Center–Crystalline Polyimide, [103320-42-9] has the following structure:



LARC-CPI has a T_g of 222°C and melts at 350°C (37). The high melting temperature illustrates one disadvantage of the high temperature thermoplastics: in order to process or melt consolidate the prepreg, it must be processed at temperatures greater than 360°C.

Poly(phenylene sulfide). Poly(phenylene sulfide) (PPS) is a semicrystalline thermoplastic with a T_g of 85°C and a T_m of 285°C. The nominal degree of crystallinity is 50–65 vol %. It is produced commercially by the direct reaction of *p*-dichlorobenzene with sodium sulfide in a polar organic solvent (38).

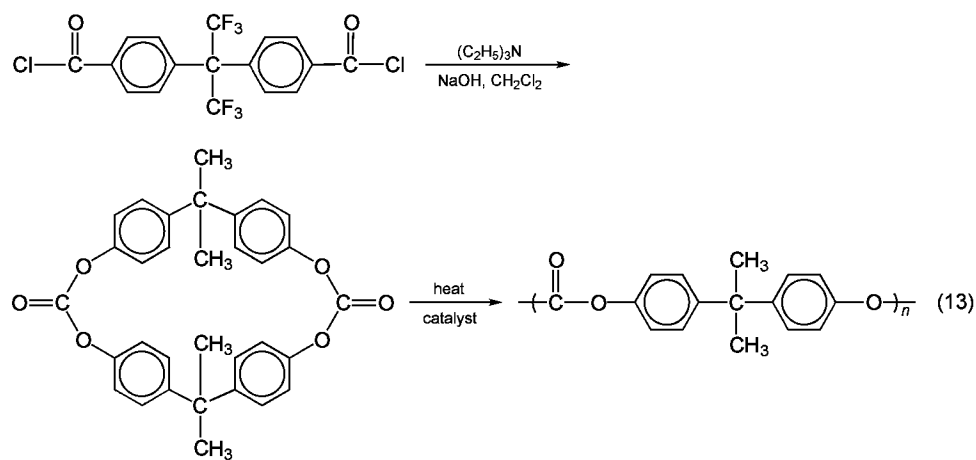


A composite utilizing the high molecular-weight extrusion grade PPS is available. Glass fiber (20–40 wt %) , carbon fiber (40–60 wt %) , or aramid (50–70 wt %) can be used as the reinforcing agent (38). Although the resin is considered a thermoplastic, continued heating above the melting point during processing produces an effect similar to that of the cure reactions in thermosetting resins and the resulting composite is pseudothermoplastic.

Novel Methods

One of the principal problems facing the use and further development of thermoplastic matrices for composites is the need to wet the reinforcing agent with the matrix to obtain a good strong bond. Amorphous thermoplastics must be heated above their T_g and crystalline materials must be heated above their crystalline melting points to flow. The upper temperature limit to which the polymer can be heated is controlled by rate of thermal degradation; thus, a relatively narrow processing window is imposed by the need to flow without degrading. In the usable temperature region, most thermoplastics are quite viscous (several hundred Pa·s) and high pressures are required for the resin to flow. Even then, in many cases the resin wets the fiber poorly. Recently, new methods have been described to circumvent this problem and still prepare a true thermoplastic composite. In one case relatively low molecular-weight, low viscosity cyclic compounds are injected directly onto preforms, the resin transfer molding process (RTM), and the polymerization conducted rapidly *in situ*. This requires a special low viscosity monomer coupled with a unique catalytic system that rapidly polymerizes the monomer but does not lead to excess thermal degradation at elevated temperatures.

The general concept of using cyclical monomers to produce high molecular-weight polymers is not new; silicones have been produced by this method for many years. However, most of the silicones are elastomers, and the extension of this method to structural thermoplastic composites is relatively new. This method (14) has been used at the General Electric Co. to prepare cyclooligomers of polycarbonate, which are subsequently used to form a composite via RTM. A mixture of cyclic carbonate oligomers was prepared by the triethylamine-catalyzed hydrolysis–condensation reaction of bisphenol A (bis-chloroformate) using methylene chloride and NaOH (15):



When the catalyst is triethylamine, the yield is nearly 100% cyclic oligomers; but if pyridine is used, the polymer is nearly 100% linear. A basic catalyst in the second step, such as lithium stearate or an organic titanate [bis-(acetylacetonato)diisopropoxytitanium], produces a polycarbonate with a molecular weight of 250,000–300,000 when polymerized at 300°C for 30 min. A fiber glass composite has been prepared using this basic procedure (39).

Another method that has great potential for the preparation of advanced prepreps and has been explored extensively requires fine powders. The reinforcing fibers are coated with fine particles of the resin and, when heated, the resin flows over the fiber. This method requires finely divided particles either in aqueous solution, in an inert volatile solvent, or as high dielectric material that can be charged and coated by electrostatic attraction to the fiber. Synthetic methods that make fine particles, similar to that described for PEEK (23), are needed.

BIBLIOGRAPHY

1. "Thermoplastic Resins," in *Engineered Materials Handbook: Engineering Plastics*, Vol. 2, ASM International, Metals Park, Ohio, 1988, pp. 98–221.
2. "Collected Papers of W. H. Carothers," *High Polymer Series*, Vol. 1, Interscience Publishers, New York, 1940.
3. M. I. Kahan, ed., *Nylon Plastics*, John Wiley & Sons, Inc., New York, 1973.
4. **U.S. Pat. 3,009,900** (Nov. 21, 1961), J. E. Hansen (to E.I. du Pont de Nemours & Co., Inc.).
5. G. Natta and P. Chini, *J. Am. Chem. Soc.* **77**, 1708 (1955).
6. K. Ziegler and H. Martin, *Angew. Chem.* **67**, 541 (1955).
7. C. A. Brighton, G. Pritchard, and G. A. Skinner, in *Styrene Polymers: Technology and Environmental Aspects*, Applied Science, New York, 1979.
8. **U.S. Pat. 2,768,994** (Oct. 30, 1961), R. N. MacDonald (to E.I. du Pont de Nemours Co., Inc.).
9. **U.S. Pat. 2,998,409** (Aug. 29, 1961), S. DalNogare and J. O. Punderson (to E.I. du Pont de Nemours & Co., Inc.).
10. **U.S. Pat. 3,027,352** (Mar. 1962), C. Welling, F. Brown, and K. Bartz (to Celanese Corp.).
11. S. M. Lee, ed., "Acetal Resins and Composites" in *International Encyclopedia of Composites*, Vol. 1, VCH Publishers, New York, 1990, p. 1.
12. W. F. Christopher and D. W. Fox, *Polycarbonates*, Reinhold Publishing Corp., New York, 1962.
13. H. Schnell, *Chemistry and Physics of Polycarbonates*, Wiley-Interscience, New York, 1964.
14. D. J. Brunelle, E. P. Boden, and T. E. Shannon, *J. Am. Chem. Soc.* **112**, 2399 (1990).
15. D. J. Brunelle and T. G. Shannon, *Macromol.* **24**, 3035 (1991).
16. W. Griehl and S. Schnock, *J. Polym. Sci.* **30**, 413 (1958).
17. E. J. Degrup, in R. B. Seymour and E. S. Kirshenbaum, eds., *High Performance Polymers: Their Origin and Development*, Elsevier, New York, 1986, p. 81.
18. W. Griehl and S. Schnock, *J. Polym. Sci.* **30**, 20 (1958).
19. C. S. Hughes, J. Bakker, and K. Tsuchiya, *Chemical Economics Handbook*, SRI International, Menlo Park, Calif. 1984, pp. 580.1172 D,F.
20. T. E. Attwood and co-workers, *Polymer* **22**, 1096 (1981).
21. **Brit. Pat. 1,414,401 and 1,414,402** (1975), J. B. Rose (to Imperial Chemical Industries).
22. C. N. Velisaris and J. C. Sefaris, *Sci. Eng. Comp. Materials* **1**(1), 13 (1988).
23. K. R. Lyon, A. Texier, A. Gangor, R. M. Davis, and J. E. McGrath, *SAMPE J.* **37**, 1301 (1992).
24. S. H. Olsen, *SAMPE J.* **26**, 31 (1990).
25. A. S. Hay, H. S. Blanchard, G. F. Endres, and J. Eustance, *J. Am. Chem. Soc.* **81**, 6335 (1959).
26. **U.S. Pat. 4,108,837** (Aug. 28, 1978), A. G. Farnham and R. N. Johnson (to Union Carbide Corp.).
27. D. Wilson, H. D. Stenzenberger, and P. M. Hergenrother, *Polyimides—Chemistry and Applications*, Blackie and Sons Ltd., Glasgow, 1990.
28. D. A. Scola, "Polyimide Resins," in *Engineered Materials Handbook*, Vol. 1, *Composites*, ASM International, Metals Park, Ohio, 1987, p. 78.
29. P. M. Hergenrother, "Polyimides Condensation" in S. M. Lee, ed., *International Encyclopedia of Composites*, Vol. 4, VCH Publishers, New York, 1990, p. 180.
30. C. E. Sroog and co-workers, *J. Polym. Sci.* **A3**(4), 1373 (1965).
31. T. Takekoshi and co-workers, *J. Polym. Sci. Polym. Chem. Ed.* **18**, 3069 (1980).
32. R. J. Boyce, T. P. Gannett, H. H. Gibbs, and A. R. Wedgewood, *Sci. Adv. Math. Proc. Eng. Ser.* **32**, 169 (1987); **U.S. Pat. 4,485,140** (1984), T. P. Gannett, H. H. Gibbs, and R. J. Kassal (to E.I. du Pont de Nemours & Co., Inc.).
33. E. Helmes, E. M. Connolly, L. Waddams, and J. Shimozata, *Chemical Economics Handbook*, SRI International, Menlo Park, Calif., 1984, pp. 580, 1401.
34. F. J. Fitzpatrick, in *Engineered Materials Handbook*, Vol. 2, *Engineering Plastics*, ASM International, Metals Park, Ohio, 1988, pp. 128–137.
35. H. Vogel and C. S. Marvel, *J. Polym. Sci.* **50**, 511 (1961).
36. **U.S. Pat. 3,822,202** (July 2, 1974), H. H. Hoehn (to E.I. du Pont de Nemours & Co., Inc.).
37. P. M. Hergenrother, N. T. Wakelyn, and S. J. Havens, *J. Polym. Sci. Part A Polym. Chem.* **25**, 1093 (1987).
38. **U.S. Pat. 3,354,129** (Nov. 21, 1967), J. T. Edmonds, Jr. and H. W. Hill, Jr. (to Phillips Petroleum Co.).
39. A. J. Salem, K. R. Stewart, S. K. Gifford, and A. M. Berenbaum, *SAMPE J.* **27**(1), 17 (1991).

Clarence J. Wolf
Michigan Molecular Institute

CERAMIC–MATRIX

With the development of emerging technologies such as aero-space, transportation and power generation, advanced materials are needed for components such as control surfaces, wing edges and nose cones; turbine blades and shrouds in more fuel efficient engines and heat exchanger elements. These structural components must operate in the temperature range between 1100 and 1650°C. Ceramics, inorganic, nonmetallic, crystalline compounds with mixed ionic–covalent nature to their chemical bonds, have been the traditional candidate materials for such high temperature use. Their many desirable properties include high melting points, high chemical stability, high elastic modulus and hardness, high wear and creep resistance, and low mass density relative to metallic materials. Monolithic ceramics, however, are brittle and are thus very sensitive to intrinsic flaws and damage produced by use. Failure of these materials occurs in a catastrophic manner and at low strain-to-failure ratios. However, the problem can be alleviated by reinforcing monolithic ceramics with a second phase which is itself capable of operating at high temperatures. Such systems are designated as ceramic matrix composites (CMC).

The reinforcing phases in ceramic matrix composites are usually also ceramic and have many possible morphologies: particulate, platelet, whisker, short-fiber, or continuous-fiber. Reinforcing entities are typically added to ceramic matrices to produce tough composites. In comparison, high strength reinforcements are added to polymer-based composites to increase strength and stiffness. To enhance toughness high strength reinforcements with high elastic modulus and weak interfaces with the matrix are required; to produce high strength and stiffness, strong interfaces along with high stress transfer are needed to allow efficient load transfer or shedding from the matrix to the reinforcement.

Ceramic Composites Systems

With the appropriate choice of composite properties, such as reinforcement and matrix materials, reinforcing geometry and composite interface, an otherwise brittle mode of failure of a ceramic becomes more "ductile" and noncatastrophic in nature. Thus, the choice of the component materials is an important aspect of designing ceramic matrix composites. Two questions need to be addressed when making these choices. First, if a matrix crack encounters a potential bridging entity, will it deflect along the reinforcement matrix interface or will fracture of the reinforcement occur? Second, if interface debonding occurs, will the interfacial sliding shear resistance, τ , be low enough to allow the bridge to slip in the matrix or will fracture of the bridging-reinforcement occur?

A partial answer to the first question has been provided by a theoretical treatment (1,2) that examines the conditions under which a matrix crack will deflect along the interface between the matrix and the reinforcement. This fracture–mechanics analysis links the condition for crack deflection to both the relative fracture resistance of the interface and the bridge and to the relative elastic mismatch between the reinforcement and the matrix. The calculations indicate that, for any elastic mismatch, interface failure will occur when the fracture resistance of the bridge is at least four times greater than that of the interface. For specific degrees of elastic mismatch, this condition can be a conservative lower estimate. This condition provides a guide for interfacial design of ceramic matrix composites.

About the second question, concerning the relative strengths of the bridge and the interfacial sliding resistance, little is known *a priori*. Some progress made for the system of continuous fiber-reinforced ceramic-matrix composites will be discussed later. The general recommendation is to have a high bridge strength and a low interfacial sliding shear resistance.

Various combinations of ceramic–matrix composites have been manufactured at the research level. Their properties are given in Table 1 for oxide-based matrices and in Table 2 for nonoxide matrices. Some commercial products are identified for information only. Such identification does not imply recommendation or endorsement by NIST, nor does it imply that the products are the best available for the purpose.

Table 1. Oxide-Based Ceramic-Matrix Composites

Reinforcement		Strength, ^b MPa	Toughness, ^c MPa√m	Modulus, ^d GPa	Density, g cm ³ or % ^e td	Reference
Type ^a	Amount, vol %					
<i>Al₂O₃ matrix</i>						
B ₄ C _p	50		4.5	380	3.28	3,4
SiC _w	20		2.5	400		5
SiC _w SiO ₂₁ ^f	20		6.0	420		6
SiC _w Si ₃ N _{4w}	20	203	3.4		95% td	7
SiO _{2f}		6.3	28			8
TiC _p	30		4.0	400	4.26	4
BN _p		24.6		490	91.1% td	9
Al _p	20		8.4			10
ZrO ₂ (t) _p ^g		2000	5–8	333	4.54	11
<i>Aluminosilicate glass matrix</i>						
SiC _w			0.8	80		5
Al ₂ O _{3f}		311	3.3			12
<i>Cordierite glass matrix</i>						
SiC _f		128	1.6		2.44	9
<i>Pyrex glass matrix</i>						
Al ₂ O _{3f}		305	3.7			12
SiC _p	30	171	1.79			13
SiC _w	30	180	3.04			13
SiC _p SiC _w ^h		159	2.73			13
<i>Soda-lime silicate matrix</i>						
SiC _w	20		0.7	72		5
<i>LASIII glass ceramic matrix</i>						

SiC _w	35	327	5.1			14
3Al ₂ O ₃ ·2SiO ₂ <i>matrix</i>						
SiC _w	10	274	2.7	197	2.84	15
SiC _p		262	2.35	240		15
ZrO ₂ (t) _p ^g		250	4.0	150	98.9% td	16
ZrO ₂ <i>matrix</i>						
ZrO ₂ (t) _p ^g		400–600	10	200	6.08	

^a Subscripts denote reinforcement morphology; p = particulate, l = platelet, w = whisker, f = fiber, i = interlayer between reinforcement and matrix.

^b Strength as measured in a four-point flexure test (modulus of rupture); to convert MPa to psi, multiply by 145.

^c Fracture toughness; to convert MPa√m to psi√in., multiply by 910.048.

^d To convert GPa to psi, multiply by 145,000.

^e %td = percentage of theoretical density.

^f 20% SiC_w.

^g Tetragonal.

^h 10% each.

Table 2. Nonoxide-Based Ceramic-Matrix Composites

Reinforcement		Strength, ^b MPa	Toughness, ^c MPa√m	Modulus GPa ^d	Density, g cm ³	Reference
Type ^a	Amount, vol %					
<i>AlN matrix</i>						
BN _p		65.5		480		9
<i>SiC matrix</i>						
SiC _w			19.9	240		8
TiB _{2p}	16		4.5	430	3.30	4
TiC _p	25		6.0	450	3.36	4
<i>Si₃N₄ matrix</i>						
SiC _w	10	620	7.8			18
SiC _w	20		4.0	350		17
SiC _w	10	436	5.7			19
Si ₃ N _{4p}		680	7.6–8.6	160		19
TiC _p		578	7.2	328		17
TiC _p	30		4.5	350	3.7	20
<i>TiN matrix</i>						
Al ₂ O _{3p} AlN ^e		229	10.2			21
<i>WC matrix</i>						
Co	20		16.9	442		10

^a Subscripts denote reinforcement morphology; p = particulate, l = platelet, w = whisker, f = fiber, i = interlayer between reinforcement and matrix.

^b Strength as measured in a four-point flexure test; to convert MPa to psi, multiply by 145.

^c Fracture toughness; to convert MPa√m to psi√in., multiply by 910.048.

^d To convert GPa to psi, multiply by 145,000.

^e 30% each.

Composite Reinforcements

The structure of reinforcements can be either equiaxed or acicular. The nature of their placement within a composite, the composite architecture, is critical to the resultant composite properties. Possible architectures are summarized in Figure 1.

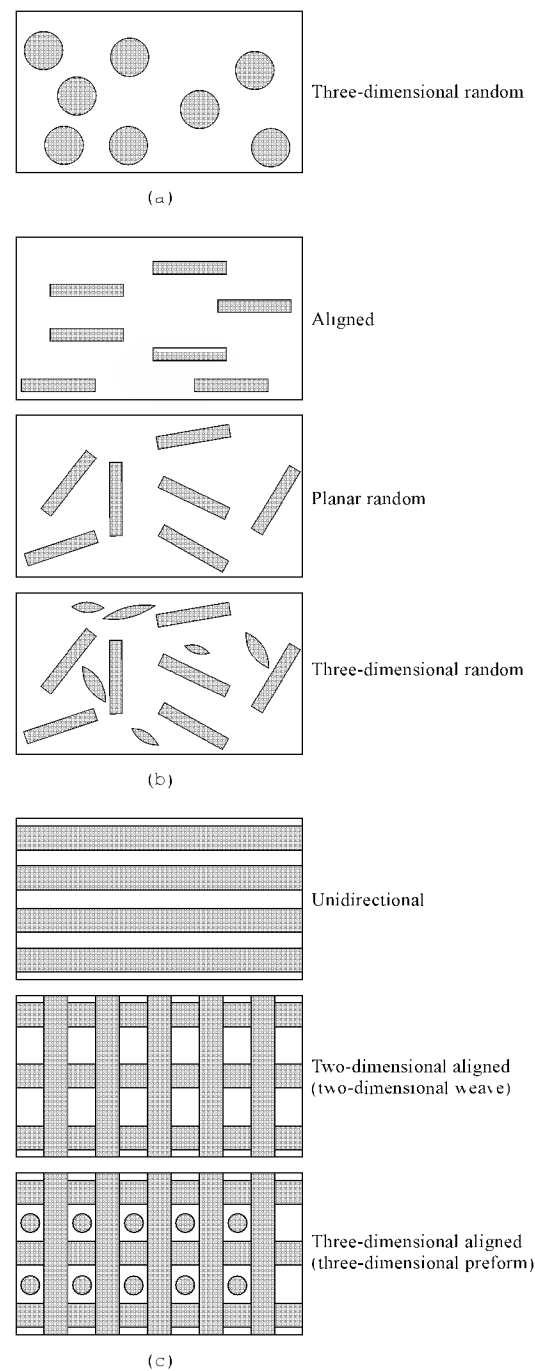


Fig. 1. Reinforcement architectures for ceramic–matrix composites and corresponding composite properties. (a) Spherical particles; (b) platelets, whiskers, short fibers; and (c) continuous fibers.

Equiaxed particles, which are well dispersed in the ceramic matrix, tend to produce isotropic composite behavior. The particles, either ceramic or metallic, may be single crystal or polycrystalline in nature.

Acicular reinforcements such as whiskers and platelets tend to produce rather more anisotropic composite properties. Whiskers and platelets are usually single crystals with aspect ratios up to 100 and with tensile strengths near their theoretical value. Composite processing can be tailored to produce either an aligned microstructure with the principal axis of all reinforcements lying in the same direction; a textured microstructure in which the principal axis is randomly arranged within a single plane; or an isotropic microstructure in which the reinforcements are randomly arranged in three dimensions. Aligned reinforcements produce a composite with highly unidirectional properties in the alignment direction, but with properties that are isotropic in the transverse direction. Such microstructures are produced when whisker-reinforced composites are fabricated by extrusion or when platelet-reinforced composites are fabricated by tape-casting or hot-pressing techniques. Textured microstructures produce composites that have isotropic properties in the reinforcement plane. This tends to be the most common type of microstructure produced when a whisker-reinforced composite is fabricated by hot pressing or tape-casting techniques. Fabrication techniques required to produce a completely random microstructure with resulting isotropic properties are extremely difficult. Hence, most ceramic composites reinforced with acicular particles tend to have some form of texture and thus, some anisotropy of properties.

Fiber reinforcements can be amorphous, single crystal or polycrystalline in structure. They can be either short fibers producing similar composite architectures to those of whiskers or they can be continuous. Continuous fiber reinforced composites tend to have orthotropic properties. For unidirectional composites properties transverse to the fiber direction are significantly different from those parallel to the fiber direction. Some continuous fibers may be woven into two-dimensional weaves before being incorporated into a composite. Such composites have in-plane properties with fourfold symmetry and different properties out of plane. These two-dimensional fiber weaves are usually rotated through a fixed angle on stacking to produce more isotropic planar properties. A preform may also be woven in three dimensions to produce more isotropic properties in all directions, but this adds considerably to the overall cost of the resulting composite.

The role of reinforcements in a ceramic-matrix composite system is to transfer stress from the matrix to the reinforcement, thereby shielding the crack tip from the applied load and providing an additional dissipative energy sink to resist crack propagation. This function is usually achieved via strong reinforcements with a weak interface between the reinforcement and the matrix. This combination allows ligament debonding and energy dissipation via frictional sliding of the reinforcement in the matrix. The matrix and reinforcement are usually chosen to allow weak interface debond stress. However, in practice, it is difficult to achieve this state because most ceramic systems react chemically. In fiber and whisker reinforced ceramic composite systems an interlayer coating of pyrolytic carbon is usually incorporated at the interface to facilitate easy debonding. Alternative approaches to weaken the interface are given later.

A related and important issue in choosing a reinforcing material is the chemical compatibility of the reinforcement with the matrix. The

reinforcement must also have high strength that is retained to elevated temperatures. If the environment has access to the reinforcement, either at the surface or through matrix cracking, then the reinforcement must be sufficiently chemically inert in the service-environment. Tables 3 through 5 present the properties of a few of the currently available platelets, whiskers, and fibers for use as reinforcements in ceramic composites.

Table 3. Platelet Reinforcements for Ceramic-Matrix Composites

Platelets	Tensile strength, GPa ^a	Modulus, GPa ^a	Density, g cm ³	Diameter, μm	Maximum use temperature, °C
Al ₂ O ₃ ^b		400	3.986	5–15 1	2040
SiC ^c	3	470	3.21	5–500 1–15	1600
SiC ^d	0.5	470	3.21	10–15	1600

^a To convert GPa to psi, multiply by 145,000.
^b Atochem, Centre de Recherche, France.
^c C-Axis, Jonquiere, Quebec.
^d Ref. 22.

Table 4. Whisker Reinforcements for Ceramic-Matrix Composites

Whiskers	Tensile strength, GPa ^a	Modulus, GPa ^a	Density, g cm ³	Diameter, μm	Length, μm	Maximum use temperature, °C
Al ₂ O ₃ ^b	20	450	3.96	4–7	40–100	2040
B ₄ C ^c	14	490	2.52			2450
SiC						
Silar SC9 ^d	7	340–690	3.2	0.6	900	1760
VLS ^e	8.3	580	3.2	4–7	5000	1400
Tokamax ^f		600	3.2	0.1–1.0	50–200	1400
SiC ^g		600	3.2	0.5–1	5–100	1400
Si ₃ N ₄						
SNWB ^h	14	385	3.18	0.05–0.5	5–100	1900

^a To convert GPa to psi, multiply by 145,000.
^b Catapal XW, Vista Chemical Co., United States.
^c Ref. 23.
^d Advanced Composite Materials Corp. (ACMC), Greer, S.C.
^e Los Alamos National Lab, Los Alamos, N. Mex.
^f Tokai Carbon Co., Japan.
^g J. M. Huber, Corp., Nacagdoches, Tex.
^h UBE Industries, Japan.

Table 5. Fiber Reinforcements for Ceramic-Matrix Composite^a

Fibers	Tensile strength, GPa ^b	Modulus, GPa ^b	Density, g cm ³	Diameter, μm	Maximum use temperature, °C
Al ₂ O ₃					
FP ^c	1.38	380	3.90	21	1316
PRD166 ^c	2.07	380	4.20	21	1400
Sumitomo ^d	1.45	190	3.9	17	1249
Safimax ^e	2.0	300	3.30	3	1250
mullite					
Nextel312 ^f	3.12	1.55	150	2.70	1204
Nextel440 ^f	4.40	2.70	186	3.05	1426
Nextel480 ^f	4.80	2.28	224	3.05	1200
SiC					
Nicalon ^g	2.62	193	2.55	10	1204
SCS ^h	2.80	280	3.05	6–10	1299
SCS6 ^h	3.92	406	3.00	142	1299
Sigma ⁱ	3.45	410	3.40	100	1259
MPDZ ^j	1.75	175	2.30	10	
HPZ ^j	2.10	140	2.35	10	
MPS ^j	1.05	175	2.60	10	
SiTiCO					
Tyrrano ^k	2.76	193	2.5	10	1300
Si ₃ N ₄					
TNSN ^l	3.3	296	2.5	10	1204

SiO ₂					
Astroquartz	3.45	69	2.2	9	993
Graphite					
T300R ^m	2.76	2.76	1.8	10	1648
T40R ^m	3.45	276	1.8	10	1648

^a Ref. 24.

^b To convert GPa to psi, multiply by 145,000.

^c E.I. du Pont de Nemours & Co. Inc., Wilmington, Del.

^d Sumitomo Chemical America, New York.

^e ICI Advanced Materials, Wilmington, Del.

^f 3M Co., St. Paul, Minn.

^g Nippon Carbon Co., Tokyo.

^h AVCO Specialty Materials Textron Inc., Lowell, Mass.

ⁱ Berghoff, Tubingen, Germany.

^j Dow-Corning Celanese, Midland, Mich.

^k UBE Industries, Japan.

^l Toa Nevyo Kogyo K. K. Tokyo.

^m Amoco Performance Products, Ridgefield, Conn.

Ceramic Matrices

Ceramic matrices are usually chosen on their merits as high temperature materials; reinforcements are added to improve their toughness, reliability, and damage tolerance. The matrix imparts protection to the reinforcements from chemical reaction with the high temperature environment. The principal concerns in choosing a matrix material are its high temperature properties, such as strength, oxidation resistance, and microstructural stability, and chemical compatibility with the reinforcement.

Another consideration is the difference in thermal expansion between the matrix and the reinforcement. Composites are usually manufactured at high temperatures. On cooling any mismatch in the thermal expansion between the reinforcement and the matrix results in residual mismatch stresses in the composite. These stresses can be either beneficial or detrimental: if they are tensile, they can aid debonding of the interface; if they are compressive, they can retard debonding, which can then lead to bridge failure (25).

Compressive interfacial stresses increase the interfacial shear resistance. Although usually detrimental to toughening, these stresses can enhance toughening if bridge pullout is the operative toughening process.

Composite Interface

The prerequisite for tough, noncatastrophic failure of ceramic-matrix composite materials is that the interface between the reinforcement and the matrix is weak enough to allow crack deflection along or around the reinforcement. For ceramic matrices and ceramic reinforcements this is not often the case as strong bonding can occur during composite fabrication. With ceramic materials primary chemical bonding occurs at the interface. One way to change the bonding is to change either the reinforcement or the matrix to materials that will not react chemically. More typically, the interface is weakened through the incorporation of a weak interlayer that is compatible with both the matrix and the reinforcements. There has been much interest in the development of weak reinforcement coatings. Alternative approaches have also been proposed; Figure 2 presents the state of the art in composite interfaces (26,27). Figure 2a represents a weak debond coating with a layer-type structure that does not react strongly with either the matrix or reinforcement, for example, pyrolytic carbon. Figure 2b presents a porous coating of the matrix material, which is much weaker owing to the presence of a large degree of porosity. Figure 2c presents a ductile interfacial layer, which has a low ductile yield stress and usually a lower elastic modulus than either the fiber or the matrix. These latter two coatings may be well-bonded to the matrix, the reinforcement, or both. These three approaches to a weak interface are discussed in more detail later.

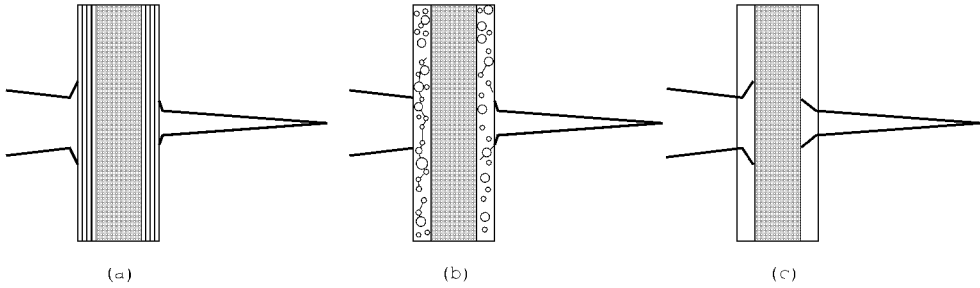


Fig. 2. Microstructural design approaches for composite interfaces: (a) mechanically weak coating; (b) porous interface and; (c) ductile interface.

A good example of the effectiveness of a weak debond coating is the Nicalon SiC–CVI SiC matrix composite fabricated at Oak Ridge National Laboratory (28) with varying amounts of pyrolytic carbon at the interface. The initial, as-manufactured, composite with no fiber coating produced a low strength, low toughness composite. Coating the fibers with a 0.17-μm-thick layer of carbon before deposition of the matrix dramatically changed the composite properties. The bend strength increased from 83 MPa (12,000 psi) to 383 MPa (55,000 psi), with an accompanying large strain-to-failure. The toughness, measured by a work of fracture, increased from 98 J m² for the uncoated composite to 4110 J m² (~2 ft lbf in.²) for the composite with the carbon-fiber coating. Figure 3 shows the corresponding load–deflection curves for the uncoated and coated composites.

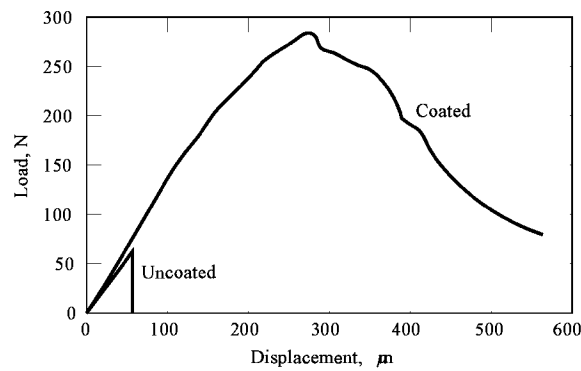


Fig. 3. Load–deflection curve for a SiC–C–SiC composite in four-point bending. Note the extreme change in behavior for a composite fabricated with a 0.17-μm carbon layer between the SiC fiber and the SiC matrix as compared with a composite with no interfacial layer (28).

Carbon is a commonly used and successful weak interfacial coating. For high temperature applications, however, carbon is not the best solution, because it oxidizes, leaving a physical gap between the reinforcement and the matrix or allowing interfacial reactions that result in a strong interface bond. Much research has been conducted to develop alternative high temperature debond coatings, with little success to date.

An alternative to the weak debond coatings is to create a mechanically weak debond interface (26). One that shows much promise is a porous coating of the matrix itself on the reinforcement (26). The coating is well-bonded to the reinforcement and the matrix but is mechanically much weaker than either because of its degree of porosity. A debond crack will thus run preferentially through the coating. Such a coating can be as temperature and oxidation resistant as the matrix itself, although the degree of oxidation protection rendered to the reinforcement phase may be reduced by any open porosity in the coating.

An alternative to the weak debond interface approach may lie in a ductile interface that is well-bonded to both the reinforcement and the matrix (27). Debonding of the interface then entails ductile yielding and shearing of the interface. Such a process potentially dissipates more energy than debonding and frictional interfacial sliding alone. The viability of such a coating has been explored for a model composite system of borosilicate glass reinforced with continuous SiC fiber (27). The coating used was electrolytically deposited copper up to 19 μm thick. The resulting composite showed an increase in toughness of 25% over the unreinforced matrix, compared with only 6% for the carbon-coated fiber composite. The disadvantage of such a system is that the degree of toughening changes with temperature, since ductility of the interlayer changes with temperature. Thus, although the high temperature toughness could be improved, the system might remain brittle at ambient temperature.

As discussed earlier, residual stresses can arise in composites when there is a mismatch between the linear thermal expansion coefficients of the reinforcement and the matrix. These stresses can be either deleterious or advantageous to the mechanical performance of the composite. Compressive stresses normal to the interface lead to enhanced interfacial sliding shear stresses. The corresponding hoop tensile stresses (and axial tensile stresses, if they arise) in the matrix reduce the strength of the matrix, and hence of the composite. Matching the thermal expansion coefficients of the reinforcement and the matrix is one solution to the problem, albeit a limited one. Another approach is the use of an interfacial layer to tailor the stresses, and the feasibility of such coatings has been proposed in a mathematical treatment of the problem (29). The results of this analysis predict that interlayer coatings with a higher expansion coefficient and a lower modulus than either the reinforcement or the matrix reduce the residual mismatch stresses. The analysis applies to the case when the thermal expansion of the matrix is greater than that of the reinforcement. The feasibility of this proposal has been demonstrated in a model composite of SiC continuous fiber reinforced borosilicate glass using a copper interlayer (27). Calculations using the theory (29) predict that a 10-μm thick copper layer will reduce the maximum hoop and radial stresses to one-third of the uncoated value. Stresses measured in the glass matrix immediately surrounding the fiber using a technique of stress-induced birefringence (30) indicate that the compressive stress of 80 MPa (11,600 psi) for the uncoated composite was reduced to a stress of less than 5 MPa (725 psi) for the copper interface composite.

Toughening Processes

The toughness induced in ceramic matrices reinforced with the various types of reinforcements, that is, particles, platelets, whiskers, or fibers, derives from two phenomena: crack deflection and crack-tip shielding. These phenomena usually operate in synergism in composite systems to give the resultant toughness and noncatastrophic mode of failure.

Crack-Resistance Behavior. The goal of composite reinforcement is to produce tough, flaw-insensitive materials that fail in a "ductile" manner. Such materials are more damage tolerant than the monolithic ceramics because they can withstand larger cracks without fracture and the fracture strength may be independent of crack size within a certain flaw size range. This important property of flaw tolerance and stable crack growth results from a fracture resistance behavior known as *R*-curve or *T*-curve behavior, in which the fracture resistance rises with crack extension. Fracture resistance can be formulated in terms of either stress intensity factor *T* or strain energy release rate *R* (or *J*). If stress intensity factor is used, then the ordinate is the square root of crack length and the plot is termed a *T*-curve. If, however, strain energy-release rate is used, *R*, (or *J*) is plotted directly as a function of crack length and the curve is termed an *R*-curve.

Increasing fracture resistance with crack length is a phenomenon common to metals tested under plane-stress conditions. Ceramics also have been shown to have similar behavior, but the phenomenon arises from elements of their microstructure that resist crack propagation (31). Ceramic composites are a natural extension of this effect, where the reinforcements provide the elements that resist fracture. The shape and extent of the *R*-curve is dependent on the microstructural scale and its effectiveness at crack shielding. If a ceramic microstructure can be designed to have an *R*-curve with an extensive initial "knee" (or portion before the inflexion point), then there is stable crack growth before failure, and hence flaw tolerance. The material is thus tolerant to flaws of a size larger than that predicted by a Griffith criterion to be unstable. The importance of toughness in engineering terms lies in the enhanced probability of detection of longer cracks via nondestructive evaluation techniques. Figure 4 is a schematic diagram of the toughness relation with crack length for a material with single-value toughness and for one with a *T*-curve. For a material with a single-value toughness, the toughness *T* can be represented by:

$$T = K_{IC} \tag{1}$$

which is not a function of crack length. For a material with a *T*-curve, the toughness can be represented by:

$$T = K_o + K_{\mu} = T(c) \tag{2}$$

In this case *K_o* is the intrinsic toughness experienced by the crack tip (equal to *K_{IC}* for a single-value toughness material) and *K_μ* is the toughness associated with the microstructural shielding term (equal to zero for a single-value toughness material).

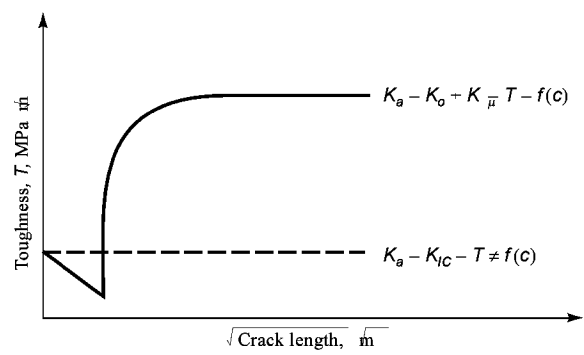


Fig. 4. Schematic representation of fracture resistance and its relation to crack length for a single-value toughness material and a material with a fracture resistance curve (*T*-curve). MPa = MN/m². To convert MPa√m to psi/in., multiply by 910.048.

The propagation of a crack depends on the shape of the *T*-curve in relation to the scale of the initial crack length. Initially, when a flaw is small, it encounters very little of the microstructure along its length. Analyses (32,33) have suggested that the initial portion of the *T*-curve decreases as illustrated in Figure 4 owing to compressive microstructural elements. As the flaw extends, more of the microstructure is sampled and the crack becomes increasingly more difficult to propagate as the toughening contribution K_{μ} increases. Once the crack is long with respect to the shielding elements in the microstructure, the toughness saturates out and the crack grows as if it is sampling an average microstructure. The equilibrium condition for fracture is that the driving force is greater than or equal to the fracture resistance:

$$K_a \geq T \tag{3}$$

The stability condition,

$$\frac{dK_a}{dc} > \frac{dT}{dc} \tag{4}$$

determines the nature of the fracture, that is, either stable or unstable. Equation 4 is the condition for unstable fracture. At equality this is known as the tangency condition.

For a single-value toughness material, $\frac{dT}{dc} = 0$. Accordingly, if the applied stress intensity factor is always increasing with crack length, equation 4 is always satisfied. Thus, the condition for fracture is equation 5, where K_a is given by the applied loading conditions.

$$K_a \geq K_{IC} \tag{5}$$

For a nonsingle-value toughness material the equilibrium condition for fracture, equation 3, becomes

$$K_a \geq T = K_o + K_{\mu} \tag{6}$$

The stability of crack extension in such materials depends on the rate of change of the applied driving force to that of the fracture resistance, equation 4.

$$\frac{dK_a}{dc} > \frac{dK_{\mu}}{dc} \tag{7}$$

Consider the toughness curve of Figure 5a. A preexisting flaw of length c_i will not extend until $K_a \geq K_o$. The dashed line represents a loading level that just satisfies the equilibrium condition for fracture, that is, equation 6. At this value of K_a , the crack propagates unstably until it arrests at a length c' ($c' > d$, where d is the length to the inflexion point on this simplistic *T*-curve). The arrest criterion is assumed to be the same as the propagation criterion, that is, equation 6, but alternative criteria have occasionally been used in the literature. Increasing K_a further, the crack propagates stably from c' to c^* , where the tangency condition, equation 7, is satisfied (see Fig. 5b) and spontaneous failure occurs. When the initial crack length $c_i > d$, there is no initial unstable crack growth, but rather the crack grows in a stable manner with increasing K_a until the crack length is equal to c^* . Preexisting flaws in the microstructure of size less than c^* but greater than a crack length c_o , as defined in Figure 5b behave in a flaw tolerant manner. The fracture strength in this initial flaw size regime is independent of crack length, as illustrated in Figure 5c. Theoretically, cracks smaller than c_o propagate unstably to failure in a high strength region. However, in practice this region is not observed experimentally because the smallest intrinsic flaws are typically of the order of d . Thus, a *T*-curve can produce a region of flaw sizes in which the material is flaw tolerant (3,34).

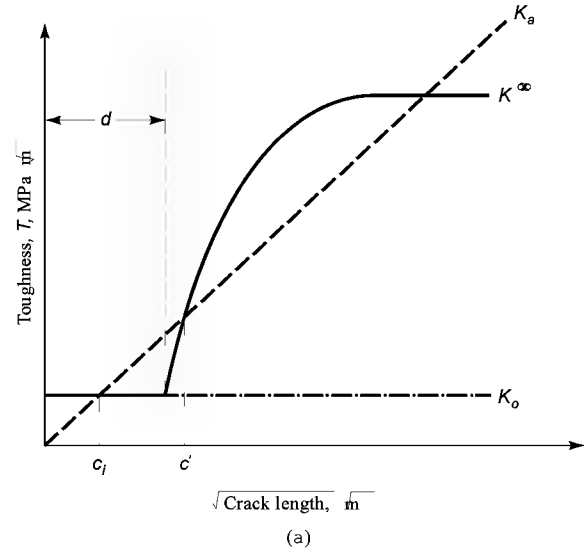


Fig. 5a. Relation of *T*-curve to crack growth and strength: (a) crack growth for a flaw of size *c_i*

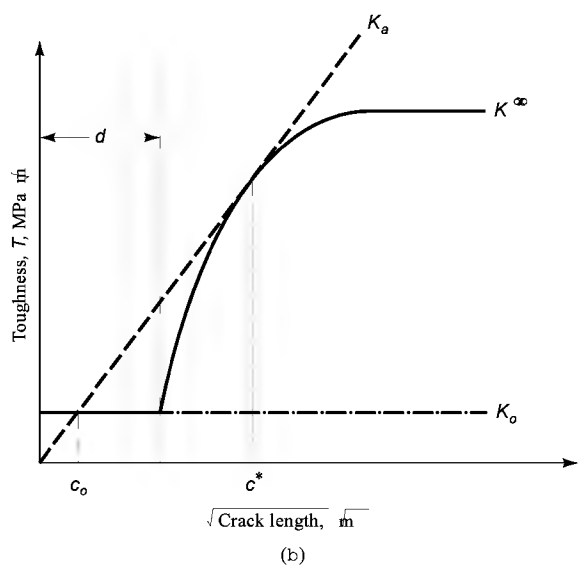


Fig. 5b. Relation of *T*-curve to crack growth and strength: (b) tangency condition for crack growth for a flaw of initial size *c_i*

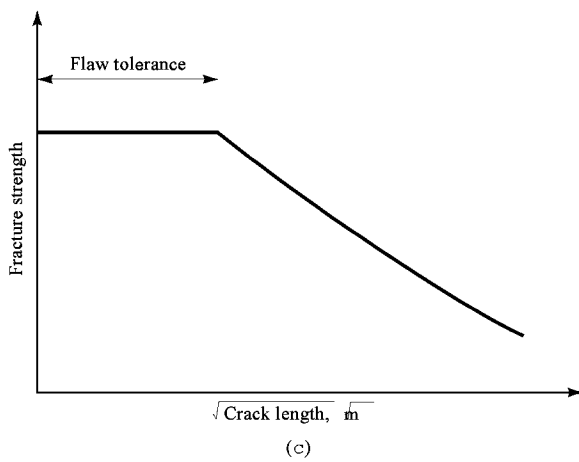


Fig. 5c. Relation of *T*-curve to crack growth and strength: (c) fracture strength as a function of crack size showing region of flaw size tolerance.

Crack Deflection Contribution to Toughening. Crack deflection is a phenomenon that leads both to toughening and to the formation of bridges that shield the crack tip from the applied stress. Little is known of the bridge formation process, but its effect, that is, crack-tip shielding, is considered in the following section.

The condition for propagation of a mode I edge crack, that is, a crack that is subjected to pure opening (tensile) stresses applied perpendicular to the crack plane, is given by (35):

$$K_a = Y \sigma_a \sqrt{c} = K_{IC} \tag{8}$$

where Y is a dimensionless geometry term, σ_a is the applied stress, and c is the crack length. Once a crack is deflected from its original plane, further crack extension requires a higher driving force to accommodate the mode II (shear) or mode III (tearing shear) contribution to the stress intensity factor on the new crack plane. A schematic diagram of crack deflection with contributions from both modes is shown in Figure 6 (36). It has been shown (37) that the net applied stress intensity factor to drive a crack on a tilt plane of θ degrees from the original plane is given by equation 9.

$$K(\theta) = K_{IC} \sec^2(\theta/2) \tag{9}$$

Similarly, if a crack is deflected from its original plane through a twist angle of ψ degrees, a mode III tearing component must also be taken into account for further propagation of the crack. The resulting toughness is given by equation 10 (37).

$$K(\psi) = K_{IC} \sec^2(\psi) \tag{10}$$

Figure 7 shows these results schematically for both twist and tilt crack deflections. Thus, for the stress intensity factor required to drive a crack at a tilt or twist angle, the applied driving force must be increased over and above that required to propagate the crack under pure mode I loading conditions. Twist deflection out of plane is a more effective toughening mechanism than a simple tilt deflection out of plane.

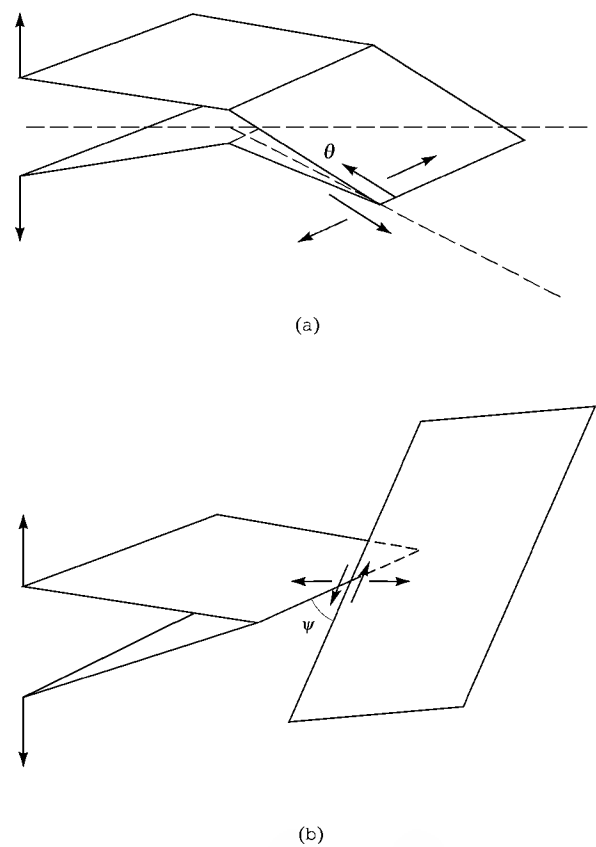


Fig. 6. (a) Crack deflection and propagation through a tilt angle $K(\theta) = K_{IC} \sec^2 (\theta/2)$. (b) Crack deflection and propagation through a twist angle, $K(\psi) = K_{IC} \sec^2 (\psi/2)$ (36).

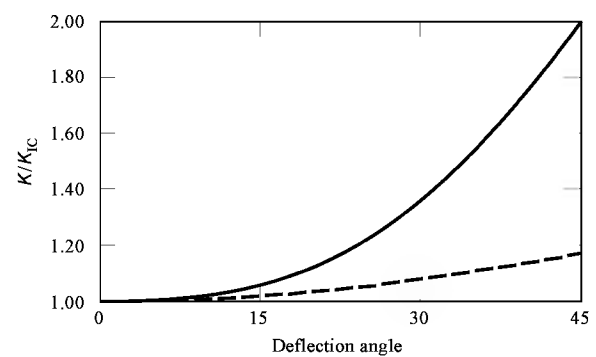


Fig. 7. Relative toughening owing to (---) tilt, θ , and (—) twist deflections, ψ .

A crack front may interact many times with inhomogeneities in the microstructure, causing multiple deflections. The resultant average driving force for a crack front that has undergone multiple deflection is an average of the many deflections of the crack tip. This problem has been treated mathematically (38–40). Having considered the effectiveness of various deflection morphologies for microstructural inhomogeneities, the authors came to the conclusion that the resultant toughening from deflection is dependent only on the shape and distribution of the inhomogeneities and not their size. The higher the density and aspect ratio of inhomogeneities, the higher the maximum toughening. Therefore, rodlike geometries produce higher toughening than either platelets or spheres. It was postulated that the toughening arises mainly from the twist component and the maximum possible toughening predicted by the model is approximately $2K_{IC}$.

Deflection rarely operates as the sole toughening mechanism in a system, although its contribution in some systems may be significant. Crack deflection, however, is a major aspect of bridge formation processes that leads to toughening via bridging ligaments.

Crack-Tip Shielding. Crack-tip shielding has two origins: process-zone shielding and crack-wake bridging. Process-zone shielding derives from mechanisms occurring in a zone around the crack tip which extend to the crack wake as the crack advances, indirectly applying closure forces to the crack flanks. Crack-wake bridging derives from intact bridging elements in the wake of the crack, directly applying closure forces to the crack flanks.

Process-Zone Shielding. An important mechanism that can lead to the phenomenon of crack-tip shielding is the development of a process zone around the crack. Stresses around the crack tip cause changes in the microstructure that can lead to process-zone shielding. In the presence of regions of nonuniform residual stresses in the microstructure, microcracks can open up, their dilation providing a back stress on the crack tip that retards further crack propagation (40). A similar dilatation can be produced by stress-induced phase transformations in which there is a volume change (41) or by yielding of ductile particles (10). Initially, a process zone is formed ahead of the crack tip, which may enhance crack growth, but as the crack advances, a fully developed process zone is formed in the wake of the crack, causing closure forces to be applied to the crack flanks. A schematic diagram presenting the identified process-zone shielding mechanisms in ceramic-matrix composites is given as Figure 8. The toughness produced by each type of zone has the same form in terms of the strain energy release rate approach (42):

$$\mathcal{R}_{\infty} = \mathcal{R}_o \left(1 + 2\eta\varepsilon V_f E'_c \sigma_c \right) \tag{11}$$

where $\mathcal{R}_{\infty}$ is the steady-state fracture resistance, (maximum in $\mathcal{R}$ -curve); $\mathcal{R}_o$ is the crack-tip fracture resistance; η is a dimensionless zone-shape coefficient (≈ 0.03 to 0.06); ε is the dilatation strain per reinforcement (microcrack, transforming particle, or ductile yielding particle); V_f is the volume fraction of reinforcement; and E'_c is the plane strain composite modulus [$E_c (1 - \nu^2)$] and σ_c the critical stress at which microcracks open up, particles transform, or particles yield.

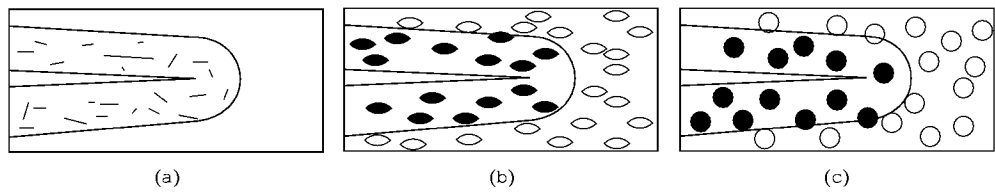


Fig. 8. Process-zone shielding mechanisms: (a) Microcrack cloud; (b) phase transformation; (c) yielding of ductile reinforcements.

The process zone toughening mechanism is seldom found to operate in isolation from other toughening mechanisms. The exception is the case of particles that undergo a phase transformation in which case the toughness is attributed to process zone shielding alone.

Crack-Wake Bridging. Crack-wake bridging occurs when matrix cracks, upon encountering a bridging entity, deflect around the entity, leaving it intact in the wake of the crack. Continued crack propagation requires further increases in the applied load to overcome the bridge closure forces. When the crack opening displacement at the bridging site is large enough to pull the bridge out of the matrix or to break the bridge, a steady state bridging zone develops, which then moves with the crack. New bridges are created at the crack front, and the bridges farthest away from the crack front become inactive. Such processes are energy dissipative and impart some degree of nonlinear stress-strain behavior to the composite; fracture becomes tougher with the possibility of developing large strains before final failure. The phenomenon of crack-wake bridging leads to *T*-curve behavior. In terms of a toughness approach, in which stress intensity factors can be linearly added, the far-field stress intensity factor K_o is equal to the stress intensity factor due to the bridging terms K_μ added to the stress intensity factor associated with the crack tip K_i :

$$K_a = K_\mu + K_o \tag{12}$$

The form of the solution for the bridging contribution is specific to the reinforcement and the crack geometry but is of the general form (35)

$$K_\mu = \sqrt{\frac{2}{\pi}} V_f \int_0^{x^*} \frac{\sigma_\mu[u(x)]}{\sqrt{x}} dx \tag{13}$$

where V_f is the volume fraction of bridging entities and $\sigma_\mu[u(x)]$ is the closure stress applied via a bridge to the flank of the crack as a function of crack opening $2[u(x)]$. The integral is evaluated over the bridging zone from the crack tip at $x = 0$ to the maximum extent of the bridging zone at $x = x^*$. Evaluation of the integral in equation 13 requires a knowledge of the crack profile with distance from the crack tip, $u = u(x)$. This function is obtained by evaluation of the integral equation (35,43):

$$2u(x) = \frac{8K_o}{E'} \sqrt{\frac{x}{2\pi}} + \left(\frac{4V_f}{\pi E'} \right) \int_0^{x^*} \sigma_\mu[u(x')] \left[2\sqrt{\frac{x}{x'}} \ln \left| \frac{\sqrt{x} + \sqrt{x'}}{\sqrt{x} - \sqrt{x'}} \right| \right] dx' \tag{14}$$

where E' is the plane strain elastic modulus. Solution of this integral equation is specific to the bridging relation and is not simple.

Toughness can also be calculated by considering a *J*-integral approach, where the equivalency with stress intensity factor is given by (44):

$$J_a = \frac{K_a^2}{E_c} \tag{15}$$

$$J_o = \frac{K_o^2}{E_m} \tag{16}$$

and

$$J_a = J_\mu + J_o \tag{17}$$

where

$$J_\mu = 2V_f \int_0^{u^*} \sigma_b(u) \cdot du \tag{18}$$

where $2u^*$ is the maximum crack opening displacement after which bridge failure occurs. Note that $2u^* = 2u(x = x^*)$. Using this approach the toughening increment, J_μ , can be calculated without having to solve the integral equation 14, however, the value obtained is only the steady-state toughening increment. No determination of the rising portion of the *R*-curve can be made using this approach. If the bridging relation includes dissipative processes, then the *J*-integral given in equation 18 is not strictly correct, because it assumes that the bridging process is nondissipative. This situation typically leads to an overestimate of the toughening increment (45).

Mechanical Performance

PARTICLE REINFORCEMENT

Particle reinforcement is an excellent method for toughening brittle ceramic matrices (3,41,46–48). The toughness imparted to such composites is due to multiple toughening mechanisms including crack deflection, crack pinning, microcracking, residual stress, frictional bridging, particle pullout, and transformation toughening. The mechanisms important to any specific system depend on the physical properties of the particles: size; morphology; thermal expansion mismatch with the matrix; and strength, toughness, and ductility.

Brittle Particles. Reinforcement via small brittle particles exploits the toughening mechanisms of crack deflection, microcracking, crack pinning, and crack bowing. The toughening contribution from the mechanisms of crack bridging and frictional pullout may be significant if the reinforcing particles are of the order of the matrix grain size or larger. All of these mechanisms arise from, or are strongly enhanced by, thermal expansion mismatch stresses in the composite. At the processing temperature the interface between the reinforcement and the matrix is stress free. On cooling to room temperature, thermal mismatch stresses can develop. If the matrix has a higher expansion coefficient than the reinforcement, the particle will be under compression and the matrix will have tensile hoop stresses and compressive radial stresses. The interface will be in compression, and the stresses in the matrix will decrease as the distance from the interface increases as a function of $1/r^3$, where r is the radial distance from the center of the particle. If, however, the particle has a higher expansion coefficient than the matrix, the nature of the residual stresses are reversed and the interface is in tension. For a particle the stresses are shown in Figure 9 (42).

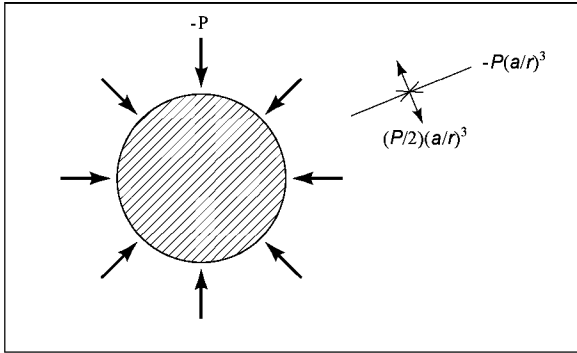


Fig. 9. Residual stresses owing to thermal expansion mismatch between a particle with radius a and thermal expansion coefficient α_p and a matrix with thermal expansion coefficient α_m . The stresses illustrated here are for $\alpha_m > \alpha_p$ and P is the interfacial pressure.

$$P = (\alpha_m - \alpha_p)(\Delta T) [(1 + \nu_m) 2E_m + (1 + 2\nu_p) E_p].$$

These residual stresses play an important role in the mechanical response of ceramic matrix composites. Compressive stresses can lead to increased strength of the reinforcement, matrix or interface. Similarly tensile stresses act deleteriously. Even though tensile stresses reduce the apparent strength of the particles, the corresponding compressive hoop stresses in the matrix can act to deflect matrix cracks, thereby leading to toughening by crack deflection. Particles in compression have tensile hoop stresses in the matrix surrounding them. These stresses can cause microcracking at the particle. When the stress field of a matrix crack combines with tensile hoop stresses, microcrack toughening is possible. In addition, tensile hoop stresses can attract matrix cracks to increase the probability of crack front pinning.

Crack pinning and the resultant crack bowing provide another toughening mechanism associated with brittle particle composites. As the driving force for the crack is increased, unpinned regions of the crack front tend to grow outward, retarded by the pinned regions. One description of the toughening that results from this phenomenon (46,49) proposes that a crack front possesses a line energy; accordingly, a crack that is pinned and bowed has a proportionally larger surface energy. The increment in fracture energy is proportional to the inverse average interparticle spacing (49) and is a function of the particle size (46). Fracture energy measurements for a composite system of alumina particles in a sodium–borosilicate glass are consistent with these functional dependencies, showing fracture energies as large as five times that of the unreinforced glass (46). (Of this increase, only a factor of less than two could be attributed to the enhanced surface roughness that results from the presence of the particles). This mechanism is not well understood and has not been widely verified; however, cracks can be pinned by inhomogeneities with a resulting tortured crack front. Figure 10 is a micrograph of crack pinning by a fiber.

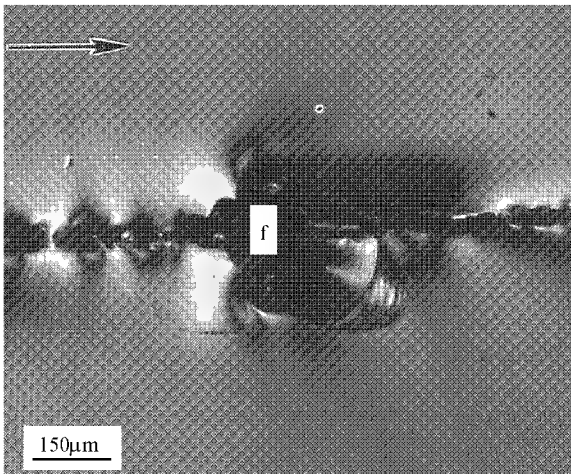


Fig. 10. Crack pinning by a SiC fiber in a glass matrix, photographed using an optical microscope and Nomarski contrast. Fiber lies perpendicular to plane of micrograph; lines represent crack position at fixed intervals of time, crack running left to right.

Courtesy of T. Palamides.

A new form of particle-reinforced composite that derives its high toughness from particle bridging and pull-out is based on $\text{Al}_2\text{O}_3\text{--Al}_2\text{TiO}_5$ (3). In this system severe thermal expansion mismatch results in residual stresses in the microstructure that enhance grain-bridging toughening. The resulting composite, with steady state toughnesses of $8 \text{ MPa}\sqrt{\text{m}}$, a higher toughness and exhibits more flaw tolerance than the matrix alumina alone. The toughening mechanism has been identified as localized grain bridging (33) and bridges, labeled B, can be clearly seen in Figure 11.

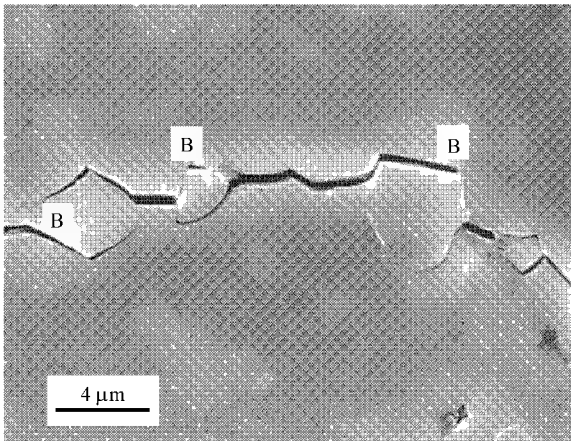


Fig. 11. $\text{Al}_2\text{O}_3\text{--Al}_2\text{TiO}_5$ typical microstructure showing grain bridging, B (SEM backscattered electron image).

Courtesy of L. Braun and S. J. Bennison.

Ductile Particles. Ductile particle reinforced ceramic composites show promise as composite material for high strength–high toughness applications. Additions of up to 20 vol % aluminum particles to a matrix of alumina have shown toughness of up to $10\text{ MPa}\sqrt{\text{m}}$ compared with typical alumina toughness of $2\text{ MPa}\sqrt{\text{m}}$ (10). Figure 12 shows the toughening mechanisms schematically.

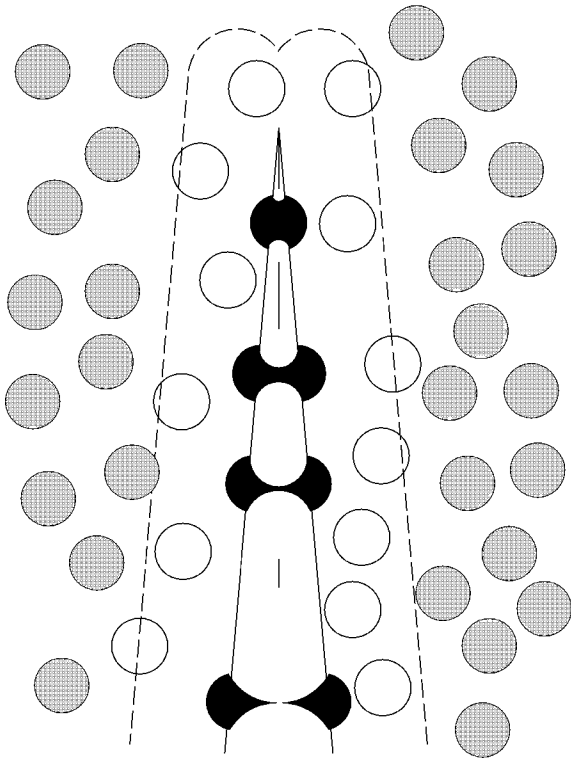


Fig. 12. Micromechanics of ductile reinforcement: particles yielding within process zone and particles bridging in crack wake.

Ductile particles can act as bridging sites in the crack wake. Instead of fracturing in a brittle manner, they undergo plastic yielding as the crack opens up (10,47,50). The maximum strain energy release rate G^∞ in the steady-state region of the $\mathcal{R}$ -curve from ductile bridging alone for a bridged-edge crack is given by equation 19 (10):

$$G_\mu^\infty = aV_f\sigma_yR\epsilon_y \tag{19}$$

where a is a geometrical constant, V_f is the volume fraction of bridging ductile particles, σ_y is the yield stress of particles in the matrix, ϵ_y is the rupture strain of the particles in the matrix and R is the particle radius. The solution for a bridged penny-shaped crack has been given (50). To enhance this mechanism large particles with a high flow stress are required.

A second toughening mechanism that operates simultaneously with crack bridging is the ductile yielding of particles in the crack-tip stress field within a process zone (10). To maximize this toughening mechanism requires a large volume fraction of particles of low yield strength.

Transforming Particles. A special type of particulate-composite are those based on the tetragonal form of zirconia. Tetragonal zirconia has the ability to undergo a stress-induced martensitic phase-transformation from its tetragonal crystal form to a monoclinic form with an accompanying dilatation of 4% unconstrained (41). Common examples of such ceramic systems include MgO or CaO partially stabilized zirconia (Mg-PSZ, Ca-PSZ, respectively), fine grain Y_2O_3 -doped zirconia (known as tetragonal zirconia polycrystals, or Y-TZP), and composites of fine-grain zirconia with alumina or mullite, (zirconia toughened alumina, or ZTA and zirconia toughened mullite, ZTM, respectively). The toughening owing to the phase-transformation-induced dilatation in terms of a stress intensity factor approach has been calculated to be that shown in equation 20,

$$K_a = K_o + K_p + T = K_o + 0.3 E e^T V_f w^{1/2} \tag{20}$$

where E is the modulus of the matrix, e^T is the transformation strain per particle, V_f is the volume fraction of transforming particles, and w is the width of the process zone (44). An equivalent solution in terms of a strain energy-release rate has been determined (51). This theoretical analysis indicates that a high modulus matrix and a large volume fraction of transforming particles are required to enhance the toughness. Such ceramic composites have been shown to have toughness as high as $15\text{ MPa}\sqrt{\text{m}}$ for Mg-PSZ with strengths of the order of 600 MPa (87,000 psi) (48). In contrast TZP composites have shown strengths as high as 2 GPa (290,000 psi) with toughness only in the range of 5 to $8\text{ MPa}\sqrt{\text{m}}$ (11). However, this particular toughening mechanism is restricted to zirconia particulate reinforcements and is only active below the phase-transition temperature, approximately 500°C.

WHISKER REINFORCEMENT

Toughening for whisker-reinforced composites has been shown to arise from two separate mechanisms: frictional bridging of intact whiskers, and pullout of fractured whiskers, both of which are crack-wake phenomena. These bridging processes are shown schematically in Figure 13. The mechanics of whisker bridging have been addressed (52). The applied stress intensity factor is given by:

$$K_a = K_o + K_\mu \quad \text{and} \quad K_\mu = K_{fb} + K_{po} \tag{21}$$

K_{fb} is the toughness associated with the stretching of partially debonding whiskers in the wake of the crack, which includes frictional sliding of the debonded region of the whisker against the matrix. K_a for the single mechanism of frictional bridging is given by equation 22,

$$K_a = \left(E_c G_o + \frac{S_w^3 R_w V_w E_c}{6 E_w \tau} \right)^{1/2}$$

(22)

where E_c is the composite Young's modulus, G_o is the strain energy-release rate associated with the crack tip, S_w is the fracture strength of the whiskers, R_w is the radius of the whiskers, V_w is the volume fraction of bridging whiskers in the crack, E_w is the whisker modulus, and τ is the interfacial shear resistance of the sliding whisker–matrix interface (τ is proportional to the friction coefficient). This solution was determined assuming a linearly increasing constitutive relationship, $P_b(u)$, for the whisker bridging zone, where

$$G_{fb} \approx 2V_f \int_0^{u^*} P_b(u) \cdot du$$

(23)

The pullout regime assumes a linearly decreasing constitutive relationship as the crack opens up, and the whiskers pull out of the matrix with increasing ease. One solution (52) for frictional pullout alone is

$$K_{po} = (I_{po} R_w) (E_c A_w \tau R_w)^{1/2}$$

(24)

where A_w is the area fraction of whiskers in the crack plane and I_{po} is the average pullout length given by

$$I_{po} \propto \left(\frac{R_w S_w}{2\tau} \right)$$

(25)

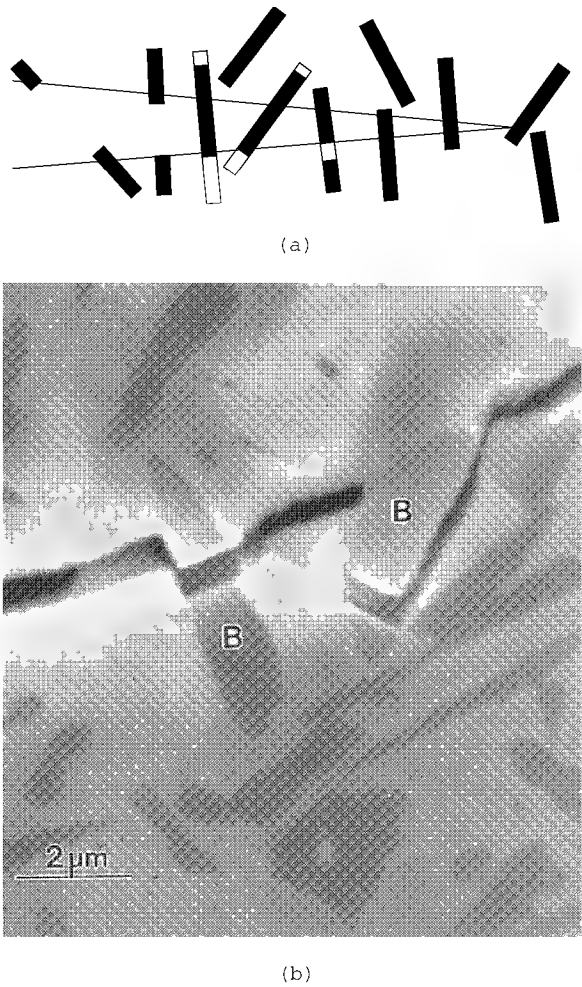


Fig. 13. Micromechanics of whisker toughening: (a) schematic diagram depicting frictional bridging, whisker fracture and pullout and (b) electron micrograph of whisker bridging in a SiC-reinforced alumina.

Experimentally it has been shown that both frictional bridging and whisker pullout play an important role in toughening industrially manufactured composites. Such investigations confirm that to maximize toughness via both mechanisms requires a high volume fraction of whiskers and a high composite modulus to whisker modulus ratio. For example, consider the effect of 20 vol % SiC whisker ($E_c = 500$ GPa) reinforcement of various matrices on the toughness as presented in Table 7 (53).

Table 7. Whisker Reinforcement of Various Ceramic Matrices^a

Matrix	Matrix modulus, ^a GPa ^b	E_c/E_w ^c	Relative toughness, K_{Ic}/K_m
glass	80	164	1.0
mullite	210	268	1.25
alumina	400	420	2.4

^a Note the importance of matrix modulus on composite properties (54).
^b To convert GPa to psi, multiply by 145,000.
^c $E_c = E_m (1 - V_w) + E_w V_w$

A high whisker strength combined with a low interfacial shear resistance τ enhances frictional bridging. Conversely, a high τ decreases the pullout contribution once whisker fracture has occurred. Increasing the whisker radius enhances both the frictional bridging and the whisker pullout contribution as evidenced by an increase in toughness of a 20 vol % SiC-reinforced alumina from 6.5 to 9 MPa $\sqrt{\text{m}}$ when the radius of the bridging whiskers used were increased from 0.3–0.75 μm to 1.5 μm (53).

Whisker reinforcement is a viable method of toughening composites. However, health considerations associated with the aspiration of fine, high-aspect-ratio whiskers raise serious concern about their widespread use.

PLATELET REINFORCEMENT

Ceramic composites reinforced with crystalline platelets show similar values of toughness as whisker-reinforced ceramic matrices. Platelets have the additional advantages of being at least one tenth the cost of whiskers, easier to process, and have higher thermal stability and none of the health hazards associated with the aspiration of whiskers. Toughness comes from a combination of crack deflection, frictional bridging and platelet pullout. Whisker toughening models are applicable if the plate thickness is aligned perpendicular to the crack front. Platelets can be of the same material as the matrix. For example, an alumina with 5 μm equiaxed grains which is reinforced with 25 vol% of alumina platelets (100–200 $\mu\text{m} \times 10 \mu\text{m}$ thick) increased the toughness from 4 to 7 MPa $\sqrt{\text{m}}$ (54). The fracture toughness increases with platelet volume fraction; eg, for the 20 vol% SiC platelet reinforced reaction bonded silicon nitride (RBSN) (55). As the volume fraction increases from 5 to 30%, the bond strength changes from 400 MPa to 350 MPa and the fracture toughness increases from 4.9 MPa to 6.5 MPa $\sqrt{\text{m}}$. Platelet size also has an important influence on toughness and strength, which generally decrease with increasing size (Table 8).

Table 8. Effect of Platelet Size and Aspect Ratio on Toughness^a

Platelet size, μm	Aspect ratio	Bend Strength, MPa ^b	Fracture toughness, MPa $\sqrt{\text{m}}$
unreinforced		600	6.0
12	1.2	610	7.9
40	4	320	6.7
70	7	300	6.0

^a For 20 vol% SiC platelet-reinforced RBSN (55).

^b To convert MPa to psi, multiply by 145.

A problem arises in using platelet reinforcements if their naturally mechanically weak crystallographic direction is aligned perpendicular to the crack front. The platelets easily fracture in this orientation. Further research is needed to grow platelets with favorable crystallographic orientations.

FIBER REINFORCEMENT

Short Random Fibers. The whiskers bridging mechanics given in equations 21 through 25 apply also to short random fiber bridging mechanisms. The bridging terms come from (44):

$$G_a = G_o + G_\mu$$

26

where

$$G_\mu = 2V_f \int_0^{u^*} P_b(u) du$$

27

The integrand, $P_b(u)$, is the force-displacement relationship for the fibers pulling out of the matrix. This relationship is identical for fibers aligned parallel to one another; however, for randomly aligned fibers this constitutive relationship is a function of fiber alignment. Studies on continuous fibers misaligned to the crack front have been conducted (56,57) to determine the constitutive relationship as a function of misalignment angle. The greatest closure forces are applied to the crack flanks when the fibers are perpendicular to the crack front. As the fiber mismatch angle increases the closure forces are reduced. These results have not been incorporated into a new bridging model.

Continuous Fibers. Composites reinforced by continuous fibers can fail in one of several possible modes depending on the interface properties and the fiber strength (58). A characterization of these modes by fiber strength and sliding shear resistance is depicted schematically in Figure 14. When the distribution of fiber strengths is broad (as characterized by a low Weibull modulus) in the regime of low fiber strength high shear resistance, fiber fracture in the crack wake occurs away from the crack mid-plane and the fibers pullout. The majority of toughening is due to the frictional pullout mechanism, although, there may be some contribution from frictional bridging before fiber fracture occurs. In this regime, the composite fracture strength is a function of the crack length, and the matrix cracking stress is the ultimate tensile stress. The mechanics associated with failure in this regime are (9):

$$G_{fb} \propto \left[R(m - 5) - \tau^{(m - 2)} \right]^{1/(m+1)}$$

28

$$G_{po} \propto \left[R(m - 3) - \tau^{(m - 1)} \right]^{1/(m+1)}$$

29

where G_{fb} and G_{po} are the toughening contributions from frictional bridging and pullout, respectively. The dependencies of these contributions on the fiber radius R and the interfacial shear resistance τ are sensitive functions of the Weibull modulus m . The toughness increases with fiber radius when m is greater than 5 and decreases when m is less than 3. The toughness increases with a decreasing interfacial shear resistance when m is greater than 2 and decreases when m is less than or equal to 1.

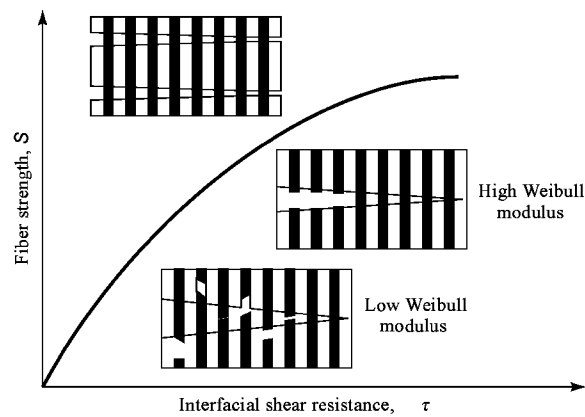


Fig. 14. Failure mechanisms for continuous fiber reinforced ceramic matrices (58).

A transition to a different mode of composite failure, which is still within this low fiber strength high shear resistance regime, occurs when the fiber strengths have a much tighter strength distribution as characterized by a high Weibull modulus. Initially the fiber strength is sufficient to allow the formation of a bridging zone in the crack wake before fiber fracture occurs at the point of highest stress in the fiber, that is, in the crack mid-plane. There is little to no fiber pullout contribution to toughening, and the contribution from fractional bridging predominates (45):

$$G_{fb} = \left(\frac{V_f V_m^2 E_m^{2m} R S^3}{3 E_f \tau E_c^2} \right) \tag{30}$$

where S is the fiber strength. Accordingly, in this low fiber strength high shear resistance regime, the fiber strength distribution, and in particular the Weibull modulus, are extremely important in tailoring composite properties to produce high toughness composites.

In the region of high fiber strength and low interfacial shear resistance, matrix cracks can propagate around the fibers leaving them intact in the wake of the crack. The matrix can be completely cracked through with the fibers supporting all the load before fiber failure begins. This regime of composite fracture leads to a very nonlinear stress-strain behavior, with a strong T -curve and crack length independent of strength at long crack lengths. Matrix cracks begin to propagate at the matrix cracking stress σ_{mu} (51,60).

$$\sigma_{mu} = \left(\frac{6 \tau G_{mc} E_f V_f^2 E_c^2}{R E_m^2 V_m} \right)^{1/3} \tag{31}$$

In such a material the toughness is primarily due to bridging contribution, rather than fiber pullout.

Chemical and Thermal Stability

Ceramic-matrix composites are a class of materials designed for structural applications at elevated temperature. The response of the composites to the environment is an extremely important issue. The desired temperature range of use for many of these composites is 0.6 to 0.8 of their processing temperature. Exposure at these temperatures will be for many thousands of hours. Therefore, the composite microstructure must be stable to both temperature and environment. Relatively few studies have been conducted on the high temperature mechanical properties and thermal and chemical stability of ceramic composite materials.

Reinforcement Integrity. Strength degradation with increasing temperature occurs to a much greater extent with ceramic reinforcements, particularly those of continuous fibers, than it does with monolithic materials. Reinforcements have high surface areas to volume so that they are more susceptible to strength degradation resulting from surface reactions with the atmosphere. These reactions can also decrease the toughness of the composite if crack-wake bridging and pullout are the predominant toughening mechanisms. This phenomenon results from the strong relationship between reinforcement strength and composite toughness predicted by the theoretical models.

Studies on the dependence of strength on temperature for Nicalon SiC fibers exposed to an oxidizing atmosphere have revealed that these fibers maintain their strength and stiffness up to 1000°C. By 1300°C, however, the tensile strength drops to 800 MPa (116,000 psi) (61). Mullite fibers (Nextel series) have a rapid fall off in elastic modulus at 900°C, but maintain a strength of at least 1 GPa (145,000 psi) up to 1400°C. Studies on strength degradation of alumina fibers (62) have revealed that these fibers maintain their strength and stiffness up to 800°C, with only 10% loss in both up to 1000°C. In general, nonoxide fibers tend to drop to half their 1000°C strength at 1400°C; oxide fibers drop to half their strength at 1100°C.

Composite Response.
Chemical Degradation. A majority of ceramic-matrix composites show strong trends in the manner in which the mechanical properties are affected by temperature. If the interface degrades, allowing strong bonding to occur between the reinforcement and the composite matrix, the toughness is considerably reduced (58). If the interface remains weak enough to allow debonding and pullout, composite strength and elastic modulus are reduced. Usually, toughness increases with increasing temperature until a temperature is reached at which failure occurs by a different mechanism, such as creep, then toughness falls off rapidly (19,63). Composites that derive their toughness from stress-induced phase transformations, such as those based on zirconia, typically undergo a drastic reduction in toughness above 800°C, owing to the change in stable phase.

If reinforcements are nonoxides, then oxidation is a possibility at elevated temperatures. Oxidation of SiC and TiC to SiO₂ and TiO₂, respectively, may be rapid at 1200°C, leading to matrix cracking as a result of the volume expansion accompanying the oxidation. If the matrix is alumina, further reactions may take place between the alumina and the oxidation products, forming mullite (3Al₂O₃·2SiO₂) with SiO₂, and aluminum titanate, with TiO₂ (Al₂TiO₅).

A further problem is possible if the reinforcements are very small. Coarsening of the particles or whiskers may occur driven by Ostwald ripening, in which large particles grow through diffusional transport at the expense of smaller ones. This can be minimized by choosing matrices in which the reinforcement elements have very low solid solubilities and diffusion coefficients. Platelets, however, have been shown to be more resistant to coarsening than particles or whiskers.

Creep Resistance. Studies on creep resistance of particulate reinforced composites seem to indicate that such composites are less creep resistant than are monolithic matrices. Silicon nitride reinforced with 40 vol % TiN has been found to have a higher creep rate and a reduced creep strength compared to that of unreinforced silicon nitride. Further reduction in properties have been observed with an increase in the volume fraction of particles and a decrease in the particle size (20). Similar results have been found for SiC particulate reinforced silicon nitride (64). Poor creep behavior has been attributed to the presence of glassy phases in the composite, and removal of these from the microstructure may improve the high temperature mechanical properties (64).

In contrast to the particulate-reinforced composites, all other reinforcement morphologies appear to provide enhanced creep resistance. The creep

rate of Ce–TZP has been found to be reduced by a factor of 5 at 1250°C by the addition of 10 vol % Al₂O₃ platelets (55). The addition of 20 vol % SiC whiskers to mullite reduces the steady-state creep rate by a factor of 10 (65).

Conclusion

Ceramic matrix composites are candidate materials for high temperature structural applications. Ceramic matrices with properties of high strength, hardness, and thermal and chemical stability coupled with low density are reinforced with ceramic second phases that impart the high toughness and damage tolerance which is required of such structural materials. The varieties of reinforcements include particles, platelets, whiskers and continuous fibers. Placement of reinforcements within the matrix determines the isotropy of the composite properties.

The toughness of ceramic matrix composites derives from a combination of two phenomena: crack deflection and crack tip shielding mechanisms. Particulate reinforced matrices derive their toughness from processes such as crack deflection, crack pinning and bowing, microcracking and frictional bridging mechanisms. Whisker, platelet, and fiber-reinforced matrices derive their toughness from crack wake frictional bridging and pullout mechanisms. For the mechanism of crack-wake bridging the most important aspect of the composite is the interface between the reinforcement and the matrix. It must be weak enough to allow debonding around the reinforcement leaving it as an intact bridging element in the crack wake. Control of the interface between the matrix and reinforcements is extremely important in optimization of composite properties.

BIBLIOGRAPHY

1. M.-Y. He and J. W. Hutchinson, *Int. J. Solids Structures* **25**, 1053–1067 (1989).
2. M.-Y. He and J. W. Hutchinson, *J. Appl. Mech.* **56**, 270–278 (1989).
3. J. L. Runyan and S. J. Bennison, *J. Eur. Ceram. Soc.* **1**, 93–99 (1991).
4. R. Warren, *Ceramic-Matrix Composites*, Blackie, Glasgow and London, 1992.
5. P. F. Becher, C.-H. Hseuh, P. Angelini, and T. N. Teigs, *J. Am. Ceram. Soc.* **71**, 1050–1061 (1988).
6. G. H. Campbell, M. Ruhle, B. Dalgleish, and A. G. Evans, *J. Am. Ceram. Soc.* **73**, (1990).
7. R. Chaim, L. Baum, and D. G. Brandon, *J. Am. Ceram. Soc.* **72**, 1636–1642 (1989).
8. F. K. Ko, *Am. Ceram. Bull.* **68**, 401–414 (1989).
9. L. M. Sheppard, *Am. Ceram. Bull.* **69**, 277 (1990).
10. L. S. Sigl, P. A. Mataga, B. J. Dagleish, R. M. McMeeking, and A. G. Evans, *Acta Metall.* **46**, 945–953 (1988).
11. K. Tsukuma, Y. Kubota, and T. Tsukidate, N. Claussen, M. Ruhle, and A. H. Heuer, eds., in *Advances in Ceramics: Science & Technology of Zirconia II*, ACerS, Columbus, Ohio, vol. 12, pp. 382–390, 1984.
12. T. Michalaske and J. Hellmann, *J. Am. Ceram. Soc.* **71**, 725–731 (1988).
13. K. W. Lee and S. W. Sheargold, *Ceram. Eng. Sci. Proceedings* **8**, 702–711 (1987).
14. K. P. Gadkaree and K. Chyung, *Bull. Am. Ceram. Soc.* **65**, 370–376 (1986).
15. M. I. Osendi, B. A. Bender, and D. L. III, (1989).
16. R. Rundgren, P. Elfving, R. Pompe, K. P. D. Lagerlnf, and B. Larsson, in S. Somiya, N. Yamamoto, and H. Yanagida, eds., *Advances in Ceramics*, ACerS, Westerville, Ohio, vol. 24B, 1988, pp. 1043–1052.
17. C. R. Blanchard and R. A. Page, *J. Am. Ceram. Soc.* **73**, 3442–3452 (1990).
18. V. J. Tennery, ACerS, Westerville, Ohio, 1989, pp. 260–272.
19. J. Homeny and L. J. Neergaard, *J. Am. Ceram. Soc.* **73**, 3493–3496 (1990).
20. C. Blanchard-Arid and R. Page, *Ceram. Eng. Sci. Proc.* **9**, 1443–1452 (1988).
21. D. Lewis, *Am. Ceram. Soc. Bull.* **67**, 1349–1356 (1988).
22. W. Boecker and co-workers, in J. Cawley, ed., *Ceramic Transactions*, vol. 2, ACerS, Westerville, Ohio, 1988, pp. 407–420.
23. A. R. Bunsell, in *Ceramic Matrix Composites*, Blackie, Glasgow and London, 1992, pp. 12–33.
24. *Ceramic Source 1991–1992*, The American Ceramic Society.
25. M.-Y. He, A. Bartlett, A. G. Evans, and J. W. Hutchinson, *J. Am. Ceram. Soc.* **74**, 767–771 (1991).
26. H. W. Carpenter and J. W. Bohlen, *Ceram. Eng. Sci. Proc.* **13**, 238–256 (1992).
27. E. P. Butler, and E. R. Fuller, Jr., "Ductile Coating for High Toughness Ceramic Matrix Composites," *unpublished work*, 1992.
28. R. A. Lowden, *4th Annual Conference on Fossil Energy Materials*, Fossil Energy AR&TD Materials Program, Oak Ridge National Laboratory, Oak Ridge, Tenn., 1990, pp. 97–113.
29. C.-H. Hsueh, P. F. Becher, and P. Angelini, *J. Am. Ceram. Soc.* **71**, 929–33 (1988).
30. E. R. Fuller, Jr., E. P. Butler, and W. C. Carter, in *NATO Advanced Research Workshop on Toughening Mechanisms in Quasi-Brittle Materials*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1990.
31. R. M. Thomson, *Solid State Physics* **39**, 1 (1986).
32. D. B. Marshall and A. G. Evans, *Mater. Forum* **11**, 304 (1988).
33. S. J. Bennison and B. R. Lawn, *Acta Metall.* **37**, 26–59 (1989).
34. Y.-W. Mai and B. R. Lawn, *Ann. Rev. Mater. Sci.* **16**, 415 (1986).
35. H. Tada, P. C. Paris, and G. R. Irwin, *The Stress Analysis of Cracks Handbook*, Del Research Corp., St. Louis, Mo., 1985.
36. S. M. Wiederhorn, *Ann. Rev. Mater. Sci.* **14**, 373–403 (1984).
37. M. Gell and E. Smith, *Acta Metall.* **15**, 253–258 (1967).
38. K. T. Faber and A. G. Evans, *Acta Metall.* **31**, 565–576 (1983).
39. K. T. Faber and A. G. Evans, *Acta Metall.* **31**, 577–584 (1983).
40. D. R. Clarke and K. T. Faber, *J. Phys. Chem. Solids* **48**, 1115–1157 (1987).
41. R. C. Garvie, R. H. J. Hannink, and R. T. Pascoe, *Nature* **258**, 703–704 (1975).
42. B. R. Lawn, *Fracture of Brittle Solids*, The Cambridge Press, Cambridge, 1992.
43. I. N. Sneddon and M. Lowengrub, *Crack Problems in the Classical Theory of Elasticity*, John Wiley & Sons, Inc., New York, 1969.
44. R. M. McMeedking and A. G. Evans, *J. Am. Ceram. Soc.* **65**, 242–246 (1982).
45. M. D. Thouless, *Acta Metall. Mater.* **37**, 2297–2304 (1989).
46. F. F. Lange, *J. Am. Ceram. Soc.* **54**, 614–620 (1971).
47. K. S. Ravichandran, *Acta Metall.* **40**, 1009–1022 (1992).
48. D. B. Marshall and J. E. Ritter, *Am. Ceram. Soc. Bull.* **66**, 309–317 (1987).
49. F. F. Lange, *J. Am. Ceram. Soc.* **22**, 983–992 (1970).
50. F. Erdogan and P. F. Joseph, *J. Am. Ceram. Soc.* **72**, 262–270 (1989).
51. D. B. Marshall, B. N. Cox, and A. G. Evans, *Acta Metall.* **33**, 2013–2021 (1985).
52. P. F. Becher, *J. Am. Ceram. Soc.* **74**, 255–269 (1992).
53. P. F. Becher, C. H. Hsueh, P. Angelini, T. N. Tiegs, *J. Am. Ceram. Soc.* **71**, 1050–1061 (1988).

54. K. B. Alexander, P. F. Becher, S. B. Walters, in the *12th International Congress for Electron Microscopy*, San Francisco Press, San Francisco, 1990, pp. 106–107.

55. N. Claussen in *11th Riso International Symposium on Metallurgy and Materials Science Structural Ceramics—Processing Microstructure and Properties*, Riso National Laboratories, Roskilde, Denmark, 1990, pp. 1–12.

56. E. P. Butler, H. Cai, and E. R. Fuller, Jr., in *Engineering Ceramics Division, 16th Annual Conference on Composites & Advanced Ceramics*, ACerS, Cocoa Beach, Fla., 1992, pp. 475–482.

57. H. Cai, K. Faber, and E. R. Fuller, Jr., *J. Am. Ceram. Soc.* **75**, 3111–3117 (1992).

58. E. Y. Luh and A. G. Evans, *J. Am. Ceram. Soc.* **70**, 466–469 (1987).

59. M. D. Thouless and A. G. Evans, *Acta Metallica* **36**, 517–522 (1988).

60. J. Aveston, G. A. Cooper, and A. Kelly, in *The Properties of Fiber Composites*, IPC Science & Technology Press, The National Physical Laboratory, 1971, pp. 15–26.

61. D. J. Pysher, K. C. Goretti, R. S. Hodder, and R. E. Tressler, *J. Am. Ceram. Soc.* **71**, 284–288 (1989).

62. A. R. Bunsell, in *International Symposium on Composite Materials & Structures*, Beijing, 1986.

63. K. M. Prewo and J. J. Brennan, *J. Mater. Sci.* **17**, 1201–1206 (1982).

64. J. M. Birtch and B. Wilshire, *J. Mat. Sci.* **13**, 2627–2636 (1978).

65. R. D. Nixon, S. Chevachoenkul, and R. F. Davis, in S. Somiya, R. F. Davis, and J. A. Pask, eds., *Mullite and Mullite Matrix Composites*, *Ceramic Transactions*, ACerS, Westerville, Ohio, vol. 6, 1990, pp. 579–604.

E. P. Butler
E. R. Fuller, Jr.
National Institute of Standards and Technology



COMPOSITES, HIGH PERFORMANCE.

See COMPOSITE MATERIALS.

COMPRESSORS.

See FLUID MECHANICS; HIGH PRESSURE TECHNOLOGY.

COMPUTER-AIDED DESIGN AND MANUFACTURING (CAD/CAM)

Computer technology (qv) has made extraordinary advances in the past two decades (1). The electronic numerical integrator and computer (ENIAC) was the first electronic computer developed for the army during World War II. ENIAC was designed to analyze artillery and bombing trajectories. It was 30 m long and 3 m high, had 18,000 vacuum tubes and contained spaghetti-like wires. ENIAC could generate enough heat to raise the room temperature to 49°C. In 1959, to eliminate wires and for miniaturization, Texas Instruments and Fairchild Semiconductors co-invented the integrated circuit in the form of small wafers called microchips. In the early 1970s Intel introduced the microprocessor, a computer on a chip the size of a thumbtack, which was capable of performing arithmetic and logical functions. Today the speed of computers has reached the rate of 100–250 million floating-point operations per second (flops), and many source-code languages have also evolved, notably FORTRAN, C, PASCAL, and BASIC. QuickBASIC, introduced by the Microsoft Corporation, is both structured and easy to use.

With the aid of computers, design processes can be expedited and implemented with greater accuracy. Computers can simulate the shape of an object, make changes, and display a three-dimensional perspective of the object on a terminal monitor. Numerous computer-aided design (CAD) software packages are commercially available. In automating chemical processes and computations, companies must decide whether to use commercial software packages or to develop in-house CAD software. For drafting needs in engineering designs, AutoCAD and VersaCAD are among the most notable commercial CAD software packages. Other software packages meet computational, word-processing, tabulation, graphing, and other needs. This article discusses many of the underlying principles that have been employed in developing popular software packages.

Except for the physical processes, such as cutting, forging, and mixing, used to manufacture items, the commonly discussed topics in computer-aided manufacturing (CAM), such as material planning, process and procedure designs, numerical control, and scheduling, might also be considered CAD because design is involved (1–3). Computer-integrated manufacturing systems (CIMS) programs have played an important role in the development of CAM activities. Flexible manufacturing systems (FMS) refers to a more recent concept of adapting basic manufacturing modules to meet a particular manufacturing need. For companies that wish to remain competitive in the international marketplace, these modern technologies in computer automation are becoming almost indispensable.

Solid modeling now plays an important role in CAD CAM. It allows a design item to be simulated on a display monitor, making possible a trial assembly of various design parts before they are actually fabricated.

Computer Application in Chemical Technology

The *Software Directory*, published by the American Institute of Chemical Engineers as a supplement to the journal *Chemical Engineering Progress*, shows the scope of computer utilization in chemical technology. Selection and evaluation of available software packages have been discussed (4). The topics covered in the second annual directory (Oct. 1990) include:

- Absorption
- Air pollution
- Automatic control
- Biochemical engineering
- CAD CAM
- Costing investment analysis
- Data acquisition
- Database management
- Data conversion
- Development tools
- Dispersion models
- Distillation
- Drafting
- Electrical engineering
- Equipment design
- Expert systems
- Extraction
- Failure analysis
- Flow analysis
- Safety
- Flow-sheet simulations
- Fluid dynamics
- Graphics
- Industrial solid waste
- Inventory
- Maintenance
- Management scheduling
- Material safety data sheets

- Mathematics
- Network optimization
- Particle dynamics
- Petroleum production
- Physical and chemical properties
- Process control
- Process economics
- Process measurement
- Programming and software
- Project and production
- Reaction kinetics
- Reactor design
- Regulations
- Reliability
- Risk analysis
- Software utilities
- Statistics
- System analysis
- Thermodynamics
- Waste management

Steady-state chemical process simulation has been reviewed (5). Uses of computers manufactured by Apple, Control Data, Gould, Telex, Lear-Siegler, MINC, ICS, IMLAC, Vector General, IBM, Megtek, and Pet for applications in process design, kinetics, heat transfer, thermodynamics, and mass transfer have been reported (6), based on two surveys conducted by CACHE (Computer Aids for Chemical Engineering) Corporation. A survey on PC-based flow-sheet software (7) compared six process engineering packages; two versions of ASPEN; CHEMCAD; DESIN 2000; HYSIM; and PROCESS.

Supercomputers, such as the CRAY X-MP, CRAY Y-MP, and CRAY-2, are partially available and used for flow-sheet and optimization studies (7–10). Optimization modules using linear and nonlinear programming (LINPRO and UNLPI, based on a revised simplex, and Davidson-Fletcher-Powell and Broyden methods, respectively) are available in MicroMENTOR (11).

Networking provides easy access to software worldwide. NSFNet (National Science Foundation network), Ethernet (Xerox Corp.), ProNet (Proteon, Inc.), and DECnet (Digital Equipment Corp.) are among the most notable networks (12).

CAD Software Packages for Chemical Processing

Many commercially available software packages, such as AutoCAD and VersaCAD, can be used to draw chemical processes in block-diagram representation. Figure 1 illustrates typical screen displays of some in-house programs. To manipulate block diagrams, numerous automated procedures are also available. Various pointer devices (mouse, light pen, joystick, and others) are used for interactive selection of particular blocks or branches for simplification and derivation of the system's transfer function. The fundamental steps that need to be taken in such manipulations, using a program called IMPROVISA, have been explained (13).

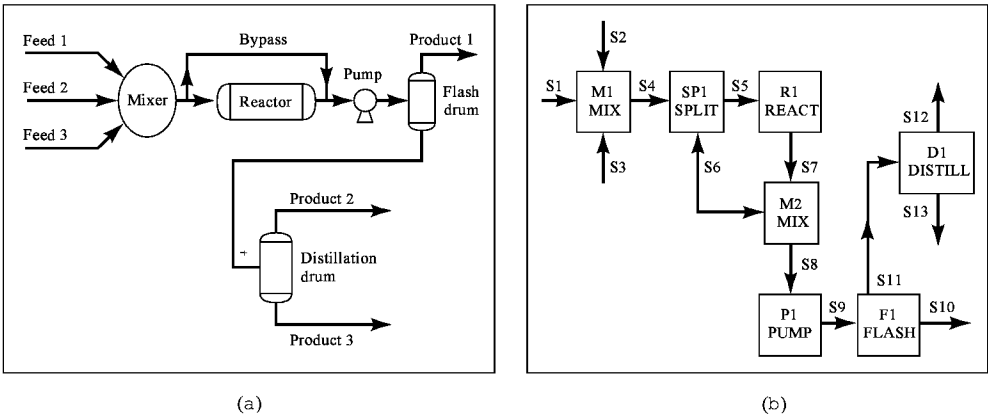


Fig. 1. Block diagrams of chemical processes.

The number of more elaborate software packages specially designed for chemical processing is rapidly increasing. Examples follow. ADVENT, developed by Union Carbide Corp., is capable of minimizing the energy and capital requirements of process synthesis. FLOWTRAN was made available in 1974 by Monsanto Co. for steady-state simulation of chemical processes based on sequential modular technology. It requires specification of feed streams and topology of the system. In 1987, an optimization enhancement was added. GPSS PC, developed by Minuteman Software (Stow, Mass.) for IBM-compatible personal computers, is based on the popular discrete simulation language GPSS (general purpose simulation system). Reference 14 provides more details. MicroMENTOR is an educational package for solving distillation problems and includes MCCABE, PONCH, and BATCH for the McCabe-Thiele, Ponchon-Savarit, and Batch binary distillations (11). The commercially available distillation software packages have been surveyed (15). For reactive distillation, ASPEN software (16) is well-known and widely adopted. SimuSolv was developed by The Dow Chemical Company for simulating physical systems and optimizing their performances. It has been applied to processing engineering, pharmacology, chemical kinetics, toxicology, environmental sciences, and agriculture. It uses DACSL (Dow advanced continuous simulation language) and applies the maximum likelihood method for optimizing the model parameters (9). TUTSIM (Twente University of Technology Simulation program) is designed for computer modeling of dynamic systems (17). It has SUM, GAIN, MULTIPLY, DIVIDE, INTEGRATE, Laplace functions, logic blocks, Z-transform blocks, and other property blocks. With a schematic design tool software OrCAD SDT III, which contains 600 library parts and over 40 integrated-circuit libraries, the TUTSIM block diagrams can be created by partitioning into small and connected, but more manageable, segments. TUT'S BLOCKS is a regular publication that promotes the use of TUTSIM. Many developers of software for finite-element analysis (18) provide drafting of pipelines and associated flow analysis. These companies include Algor, McAuto, MacNeal-Schwindler, and FlowDesign. In software, in-house developed pipe fittings are modularized and isometric views of the piping systems with three-dimensional detailing are now commonplace. Many more special-purpose software packages have been developed, particularly in teaching and research institutions. SMCM is software designed at the University of California in Los Angeles for partitioning of pollutants (19). Monte Carlo and molecular dynamic techniques have been adapted in a

molecular simulation model at Cornell University for presenting a detailed atomistic description of the material under study (20). Novo-Nordisk A/S, a Danish biotechnology company, and Silicon Graphics, Inc. have also applied similar techniques for new drug development.

Computer Graphics in Instruction

The use of graphic displays as an essential element of computer-based instructional systems has been exploited in a number of ways. Molecular modeling and visualization techniques have supplemented the traditional set of stick models in courses on organic and inorganic chemistry, and animation of molecular motion and of the progress or mechanism of chemical reactions has been a useful classroom tool.

Some practical experiences in using graphic presentation of process simulation results to teach process analysis and dynamics have been described (21). Interactive graphic displays have been included in teaching chemical reaction engineering, process control, and staged operations (22). The development of computer graphics in the PLATO system at the University of Illinois has had a significant impact on the use of computers in the chemistry and chemical engineering curricula. Experiments such as distillation are modeled by allowing the user to assemble the apparatus from pieces shown on the screen, charge the distilling flask, boil off the contents while accumulating a T - x curve, and even experience an explosion if the receiver overflows. A number of groups have explored the use of PLATO for instruction in chemical engineering (23–25).

The ubiquitousness of the personal computer and the increasing power of the program packages available for it have shifted the emphasis away from large, mainframe-based systems like PLATO. Instead, graphics are included in utility programs such as equation solvers, spreadsheet processors, and word processors.

Computer Graphics and Chemical Structural Analysis

CAD/CAM techniques have provided the framework for using the computer as a tool in the drawing and analysis of chemical structures and, more recently, in the use of chemical structures to design reaction pathways and new products. The essential elements in these applications of CAD/CAM are that the possible structures are relatively deterministic and that allowable changes in structure through reaction are governed by thermodynamic, stoichiometric, and steric constraints.

Two-Dimensional Representation of Chemical Structures. The IUPAC standardization of organic nomenclature allows automatic translation of a chemical's name into its chemical structure, or, conversely, the naming of a compound based on its structure. The chemical formula for a compound can be translated into its structure once a set of semantic rules for representation are established (26). The semantic rules and their application have been described (27,28). The inverse problem, generating correct names from chemical structures, has been addressed (28) and explored for the specific case of naming condensed benzenoid hydrocarbons (29,30).

An important approach to the graphic representation of molecules is the use of a connection table. A connection table is a data base that stores the available bond types and hybridizations for individual atoms. Using the chemical formula and the connection table, molecular structures may be generated through interactive graphics in a menu-driven environment (31–33) or by using a linear input of code words (34,35). The connection table approach may be carried to the next step, computer-aided molecular design (CAMD) (36).

A difficulty arises in representing chemical structures in print or on a CRT (cathode-ray tube) because chemical structures are three-dimensional and the representations are two-dimensional. One approach to overcoming this problem has been the use of the linear notation developed by Wiswesser (37) to represent three-dimensional structures two-dimensionally. An implementation of this method for computer applications has been proposed (38). The CHEM program for phototypesetting (39) is an example. The widespread availability of text editors for electronic publishing has led to a number of commercial programs for drawing chemical structures (40,41). These programs may be stand-alone, or they may create files that can be integrated directly with text files.

The Chemical Abstracts Service has institutionalized the use of graphic representations for identification of and retrieval of information about chemical compounds through their Graphical Data Structure (GDS) and connection table (CT). Two information packages, Messenger and STN Express, are the basis for this on-line retrieval system (42).

Three-Dimensional Modeling of Chemical Structures. The two-dimensional representations of chemical structures are necessary to depict chemical species, but have limited utility in providing true understanding of the effects of the three-dimensional molecule on properties and reactive behavior. To better describe chemical behavior, molecular modeling tools that reflect the spatial nature of a given compound are required.

Early efforts to develop molecular models emphasized ways of representing three-dimensional aspects in two-dimensional projections. Some of the problems addressed were the folding of macromolecules (43,44) and two-dimensional projections with hidden surfaces (45,46). The state of the art in the early 1970s has been reviewed (47).

A number of groups developed approaches to molecular modeling, including a graphic system (48–50), the MAGIC system (51), the DRACO program (52), and the GRAMPS/GRANNY system (53). One relatively early program explicitly designed for producing three-dimensional structural representations was CDRAFT (54). This program provides an array of structures that can be assembled using a template and a set of special symbols.

The three-dimensional models used in modern organic chemistry have their origins in Newman projections (55,56). The projections may be depicted using a stick-and-ball representation known as ORTREP (57) or a perspective-shading approach (also known as space filling) used in the SPACEFIL program (58). Ways to incorporate both stick and ball and perspective shading into molecular representations have been suggested (59). The SPACEFIL program has been modified to incorporate color graphics (60), and most of the currently available molecular modeling programs use the combined depictions.

A unique application of molecular modeling techniques has been the recreation of the Brookhaven Protein Data Bank (PDB) in the form of computer-generated molecular models (61). Many other programs use the space-filling approach (62–64). Interactive programs for macromolecular modeling include the GRAMPS/GRANNY system (53) and CHOODRAW (65).

Three-dimensional representations of complex molecules have opened the way for a number of applications. For example, structural models of opened double-stranded DNA have been used to study interactions of DNA with various intercalating drugs (66). Enzyme-inhibitor complexes have been studied using the models (67) and a movie illustrating the actions of zinc protease thermolysin that was made using the GRAMPS system (68). Graphic conformational analysis has been extended to peptide systems using the PepCAD system (69) to determine the minimum conformational energy for *N*-formyl-*N'*-methylalanineamide.

SPACEFIL has been used to study polymer dynamics caused by Brownian motion (60). In another computer animation study, a modified ORTREP II program was used to model normal molecular vibrations (70). An energy optimization technique was coupled with graphic molecular representations to produce animations demonstrating the behavior of a system as it approaches configurational equilibrium (71). In a similar animation study, the dynamic behavior of nonadiabatic transitions in the lithium–hydrogen system was modeled (72).

Molecular Modeling for Reaction Path Synthesis and Molecular Design

Applications of CAD/CAM to drawing and modeling chemical structures are essentially passive; that is, although the user decides which compound to draw, strict rules govern what that compound will look like. A reasonable goal for the application of molecular modeling capabilities is the ability to design chemical syntheses based on the structure of a desired product, a set of potential reaction pathways, perhaps defined on the basis of a connection table and a Gibbs energy minimization routine (see MOLECULAR MODELING).

Emulation of the reasoning path of a chemist in developing molecular structures was the basis for the program CONGEN (73). The reasoning was guided by a set of constraints applied to the evolving structure. However, determination of a structure on the computer must be guided by more than a comprehensive set of rules. For example, spectroscopic information must be considered (74,75).

Techniques for the synthesis of reaction paths on the basis of enumeration of all possible reactions are fairly well developed. The SYNCHEM2 system conducts a backward search of possible reaction paths from a desired product (76). Because the procedure is limited by the rapid growth of

combinatorial problems, a heuristic approach for choosing desirable paths was developed. This enumeration approach was first suggested in 1969 (77).

An alternative approach (78,79) is based on a set of possible reaction schemes that are used to generate potential new pathways. Under both approaches, the problem, in part, is how to evaluate the utility of a particular scheme. A computer-assisted approach to predicting potentially useful reactions has been developed (80). The union of existing capabilities in modeling chemical structures with selecting reaction pathways has not yet taken place.

An area that has used chemical structures for predictive purposes quite successfully is the estimation of thermophysical properties of compounds. There has been an extensive compilation of estimation methods (81), and prediction of physical properties has been automated using these techniques (82). More recently, the use of group contribution techniques to design new molecules that have specified properties has been described (83). This approach to compound design is being used to develop replacement materials for chlorofluorocarbons.

Pattern Recognition and Interactive Graphics Applications

The graphics capabilities of the CAD/CAM environment offer a number of opportunities for data manipulation, pattern recognition, and image creation. The direct application of computer graphics to the automation of graphic solution techniques, such as a McCabe-Thiele binary distillation method, or to the preparation of data plots are obvious examples. Graphic simulation has been applied to the optimization of chemical process systems as a technique for energy analysis (84).

The interactive features of the modern computer allow the use of graphic methods to be further expanded. An early effort to couple data fitting with interactive computing was the VIPER program (85). In the VIPER system the user changes parameters within a curve-fitting routine with a light pen, after viewing the results from the previous fit. A somewhat similar approach was suggested for analyzing results from a shock-tube experiment by selecting candidate kinetic models, viewing the results, and then modifying the model until satisfactory results are obtained (86). The capability of interactively selecting multiple fitting models for a set of data is now commonly available in a number of spreadsheet programs, such as Lotus 1-2-3 and Quattro-Pro. The use of Lotus 1-2-3 in multivariant kinetic analyses has been described (87).

The graphic operating framework has been effective as a basis for image processing applications. Early work (88) using digitized images of bubbles in a gas-liquid reactor demonstrated the utility of interactive image processing in research. Examples of similar studies include drop-size studies in sprays (89) and mixing studies (90).

The use of graphic simulators for flow visualization has grown from relatively simple models, such as the generation of natural convection streak lines (91), to sophisticated commercial programs for the analysis of complex flow regimes. Examples include the simulation of plastic injection molding by Graftek, Inc., and the flow simulator FLOW3D developed by the HTFS consortium. Perhaps the ultimate graphic simulation is the visual depiction of process equipment as it is being designed, using a program such as the Intergraph, Inc., Design-Review system.

CAD/CAM Applications in Process Flow-Sheet Development

The use of the computer in the design of chemical processes requires a framework for depiction and computation completely different from that of traditional CAD/CAM applications. For this reason, most practitioners use computer-aided process design to designate those approaches that are used to model the performance of individual unit operations, to compute heat and material balances, and to perform thermodynamic and transport analyses. Typical process simulators have, at their core, techniques for the management of massive arrays of data, computational engines to solve sparse matrices, and unit-operation-specific computational subroutines.

The more traditional CAD/CAM approaches do, however, have a significant place in chemical process design. The creation and optimization/modification of a process flow diagram lends itself to the CAD/CAM framework. Early work (92) was restricted to the synthesis of piping networks as components of the process flow sheet. A system specifically designed for the creation of process flow diagrams using an experimental program, PFDII, has been described (93,94). PFDII was based on an adaptation of Rubin's graph reduction and embedding procedure to construct and automatically modify the process flow sheets. In a somewhat different approach (95) a relational database was employed to store and retrieve the symbols for creating the diagram, and interactive methods allowed the user to both create and update the flow diagram.

The introduction of menu- and icon-driven interactive programs has permitted improvements in computer-user interfacing. Two programs, PFG for process flow sheets and PIG for piping and instrumentation diagrams, use icon and menu input (96). ChemShare has offered a version of the DesignII program that interfaces directly with Microsoft Windows, providing a traditional process design simulator with the capability of drawing process flow schematics.

One opportunity for the use of CAD/CAM solids visualization techniques is in the three-dimensional depiction of chemical process equipment as a possible replacement for the process models that are now commonly used as guides for construction piping and equipment layout. An interactive program (97) for this purpose uses a number of primitives generated from polyhedrons to create shape clusters that can be used to represent the three-dimensional shapes of processing equipment. An equipment visualization capability is incorporated in the Intergraph, Inc., DesignReview software.

Most of the recent flow-sheet drawing applications involve interfacing with process design simulators so that a flow sheet entered into the simulator may also be depicted by associated graphics software. One example of add-on software is a program called GRAF that accepts input files from the ASPEN simulator to create a process flow diagram (98). Some commercially available packages that include graphics as integral parts of the program are ChemcadII from COADE, ASPEN PLUS with ModelManager, and ProSim. The ChemShare Design PFD system is an example of a process flow-sheet program designed to operate in association with a comprehensive process simulator.

A future goal for the integration of graphics and process design simulators is to be able to use an interactive graphics program to prepare the input to the process simulator. This capability would allow true on-line process modification, flow-sheet optimization, and process optimization, and is likely to be one of the key developments in this field in the 1990s (99).

Graphics Applications in Process Optimization and Control

The capabilities for computer-aided process design and for utilization of CAD/CAM techniques in association with design have progressed further than similar applications in the design of process control systems. Direct application of graphic techniques to the design of process control systems has been limited to certain approaches, for example, one in which graphic display of the frequency response of feedback controllers is used as part of an interactive design method for feedback control systems (100). The CHESOPS program (101) took a similar approach to operability analysis of a process. The basis for performance calculations with the CHESOPS system was a conventional process simulator, which was run for a number of cases to simulate dynamic behavior. The graphics consisted of a series of two-dimensional plots that explored the feasible operating region for the plant. The optimization of a semibatch free-radical polymerization using interactive CAD tools (102) also used graphic display of computed results as an aid to interactive analysis.

The goal of integrating automated process optimization with automated design of process control and instrumentation diagrams will likely be more difficult to attain than the integration of graphic input for process simulation programs. The computational engines for optimization are under development at this time, as are effective dynamic simulators.

There are, however, two areas in which graphic methods have had a significant effect on both process design and process control. The synthesis of heat-exchange networks is an exercise in examination of multiple combinations of hot and cold streams transferring heat across heat-exchanger nodes. The network synthesis methods (103) are intrinsically graphic in approach. An interactive software package, RESHEX (104,105) has been developed to implement this approach. The HEXTRAN program by Simulation Sciences, Inc., is another example of the automation of heat-exchanger network design.

The second area, the implementation of a modern process monitoring and control system, is the most dramatic current application of CAD/CAM technology to the chemical process industry. The state of the art is the use of computer graphics to display the process flow diagram for sections of the process, current operating conditions, and controller-set points. The process operator can interact directly with the control algorithms through the

keyboard or with a light pen to change the set points. The controllers may consist of conventional feedback or feed-forward loops that are coupled directly with the process; alternatively, they may be devoted simulators that make decisions based on computed projected performance.

The process monitors and controllers typically also have the capability for data logging, analysis, and display. This capability has made on-line control of pilot plants, as well as commercial-scale processes, desirable. Pilot-plant applications for on-line control have been described (106), and the use of such systems for both monitoring and process diagnosis has been discussed (107). A number of commercially available process control programs that run on microprocessors have been reviewed (108). Virtually all of them incorporate graphic display as an integral part of the interactive capability of the program.

Essential Elements for Developing In-house, Special-Purpose CAD Software

Any program developed in-house must be easy to use, or user-friendly. If the program has various options for input, analysis, computation, and output, then it must provide the user with a fast way to select them. To meet this need, the system is likely to be menu-driven. The peripheral interactive devices such as mice, joysticks, light pens, graphic tablets, and templates are helpful and often used to expedite the selection process.

User-Created Cursors of Various Shapes and Sizes.

When a microcomputer system is powered on, the connected display monitor shows a flashing _ (underline), a filled rectangular block, or another predesignated shape. Such prompting signs are cursors that alert the user to enter character or characters via the keyboard as part of a program or as a command. Figure 2 provides an example of two types of cursors used in the AutoCAD software. The left panel shows a filled bar cursor appearing as the background for the word "DRAW," which can be moved up and down by pressing the respective arrow keys on the keyboard; the right panel shows two hairline cursors, one horizontal and one vertical, which are moved using a mouse.

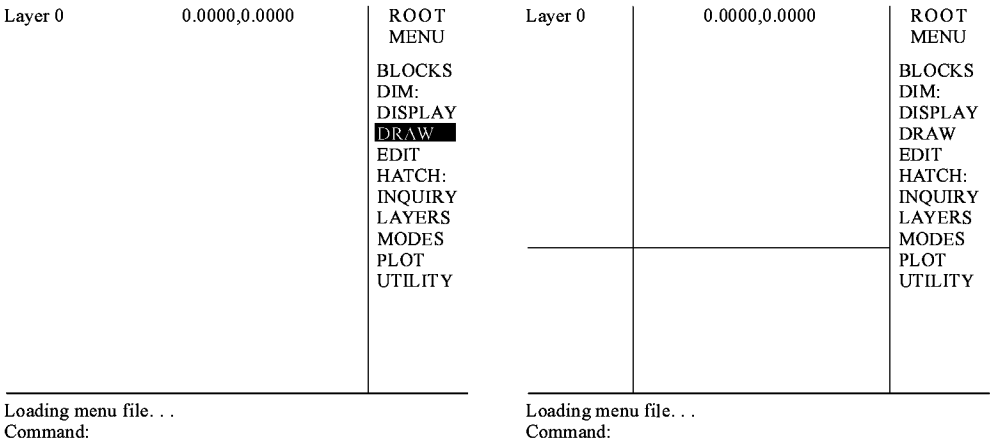


Fig. 2. Rectangular cursor (highlighted) in the left panel and horizontal and vertical hairline cursors in the right panel.

Courtesy of Autodesk.

Cursors of different shapes and sizes can be designed easily by the microcomputer user through simple programming, such as using the BASIC commands GET and PUT. The movement of the user-created cursor is controlled through the use of the INKEY\$ function and by testing the ASCII codes of the keyboard keys.

Two unfilled rectangular bar cursors of the same size and shape are used in Figure 3. The top bar cursor allows a particular menu to be selected by pressing the left or right arrow keys, and the lower bar cursor allows one of the items listed under the chosen menu to be selected by pressing the up or down arrow keys. A mouse can also be used to control the selection.

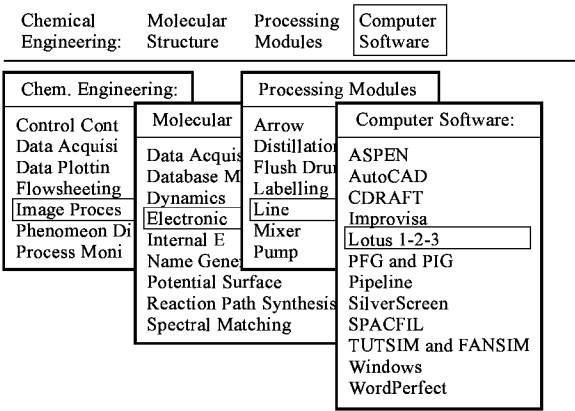


Fig. 3. Nested menus.

Menu, Tablet, Template, and Spreadsheet.

Menu design is an essential element in microcomputer software development. Figure 2 shows a menu, "ROOT MENU." Figure 3 is presented to show all available menus. When there are more menus and the total width exceeds 80 columns, scrolling must be effected (109,110). If the top bar cursor is moved to the first column, Menu 1, only the contents of Menu 1 will be displayed.

A tablet is a device for placing selectable menus or modules on a separate, peripheral plate. It is helpful for a user of CAD software to have all the modules available displayed on the screen for selection. The only drawback in such an arrangement is that it takes away a portion of the screen space and therefore less space is available for the actual CAD activities. Selection of an item from a tablet can be made with a mouse or light pen (Fig. 4).

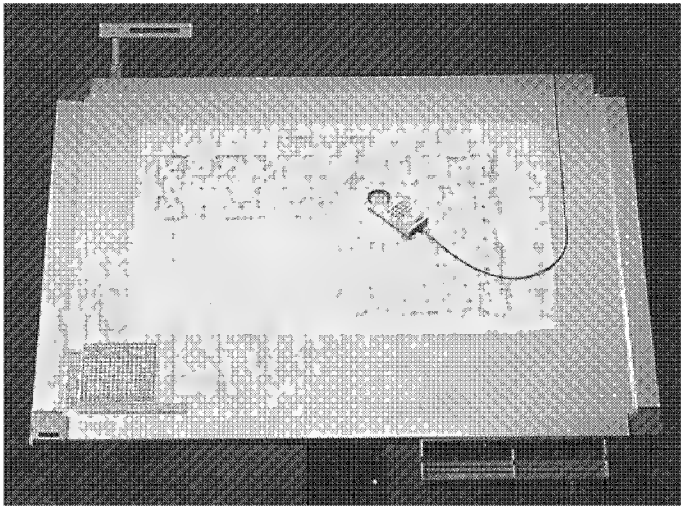


Fig. 4. Selection from a tablet with a light pen. Courtesy of CalComp Inc.

A template can be designed to be mounted or pasted on a keyboard for quick reference. It shows which function keys, in conjunction with the Shift, Alt, and Ctrl keys, have been assigned for implementing special tasks. A well-known example is the WordPerfect template shown in the left panel of Figure 5. A template is usually designed in conjunction with the function keys available on the keyboard. Template.Fon is a program developed to interactively use the display screen to design templates (1). In the right panel of Figure 5, specially designed fonts stored as disk files are drawn at desired locations to indicate the assignments of the 40 function keys for displaying, in this case, elements involved in chemical technology.

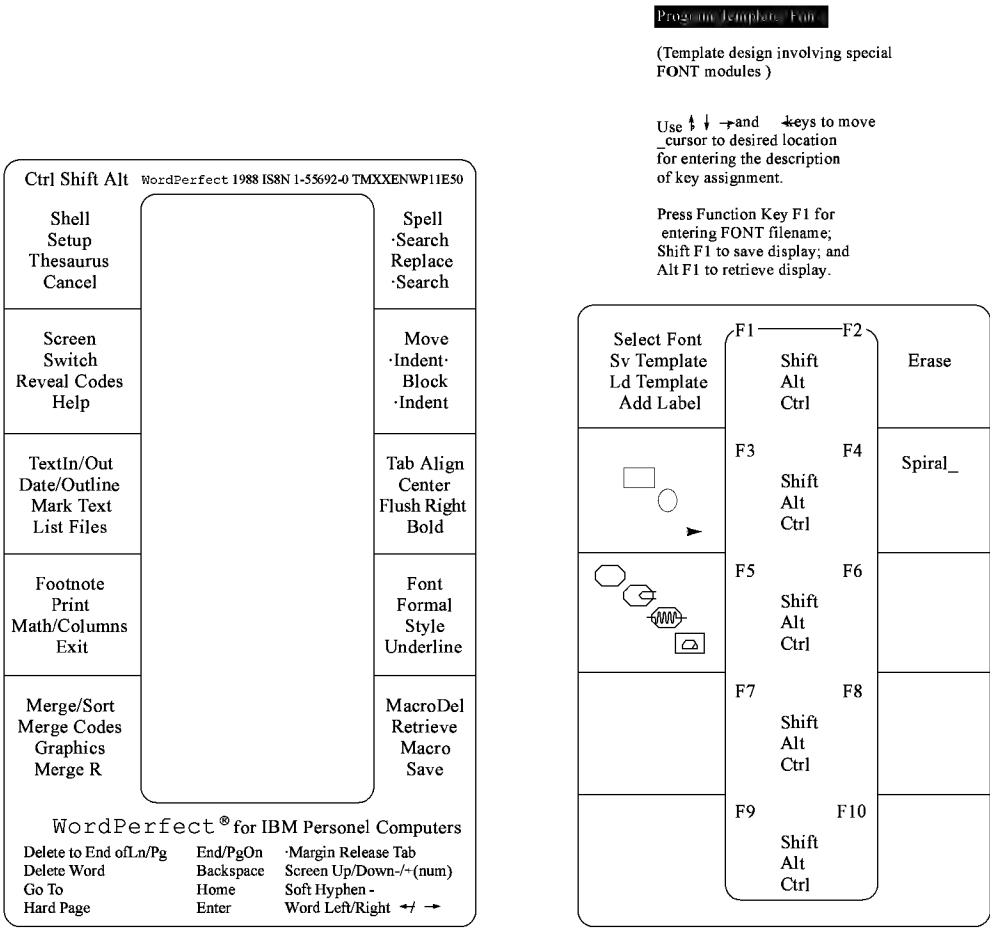


Fig. 5. CAD template. Left panel, courtesy of WordPerfect Corporation.

A spreadsheet is another essential element in developing CAD software. LOTUS 1-2-3 is well-known among the numerous commercially available spreadsheet software packages. Spreadsheet software is used to design tabulated forms for storing and editing data, for listing programs in matrix form, and for easily combining various elements in a matrix to print or to generate new rows and columns. Quick recall and access to a data set or program and the ability to combine it with other data sets or programs are the principal goals of designing spreadsheets. A spreadsheet can be used for drawing flow sheets and for optimizing chemical processes (111).

Modular Approach and Three-Dimensional Solid Modeling

Commonly used entities such as the pipe fittings shown in Figure 6 are increasingly being created and stored as graphic files for quick retrieval. In the construction of three-dimensional piping systems, they can be retrieved, rescaled, and positioned by use of a pointing device for connecting to the other pipe elements. Once a three-dimensional pipeline system has been constructed, it can be displayed on the screen using multiple views. Figure 7 is a four-way, split-screen display showing the front, top, isometric, and right-side views of an incomplete pipeline system. Detailing of a three-dimensional structure can be carried out in a multidimensional way, as illustrated in Figure 8. SilverScreen, developed by the Schroff Development Corp., is an example of commercially available software that is capable of implementing the latest trends in computer-aided drafting. It provides the capability for detailed annotation on the front, top, and right plane of a three-dimensional solid model, as shown in Figure 8. It is menu-driven and divides the screen, as shown in Figure 9, into a menu area (upper left corner), a status area (lower left corner), and a graphic area (the remaining portion on the right). SilverScreen

enables the graphic area to be split up to eight ways for displaying a three-dimensional object. A pointing device such as a mouse or the arrow keys on the keyboard can be used to select an option from the menu list, and the Del (delete) and Ins (insert) keys on the keyboard can be pressed to adjust the increment being used for the movement of the screen cursor. The three-dimensional coordinates of the current position of the screen cursor (x, box, or trid) and the increment of its movement on the screen are all displayed in the status area. In Figure 9, the Menu Area shows a variety of the capabilities of SilverScreen. For example, the Language option enables computer programs written in BASIC and C to be implemented in conjunction with the SilverScreen commands. The I/O option enables the drafting or solid model files created by SilverScreen to be stored in a variety of formats compatible with other CAD/CAM software packages. The Window option enables a three-dimensional solid to be viewed using up to eight windows. The Presentation option enables a three-dimensional object to be displayed with the hidden lines removed and shaded with different grays or rendered in color.

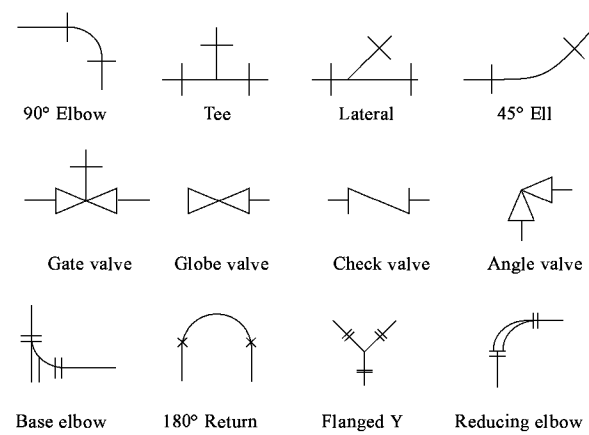


Fig. 6. Pipe fitting elements.

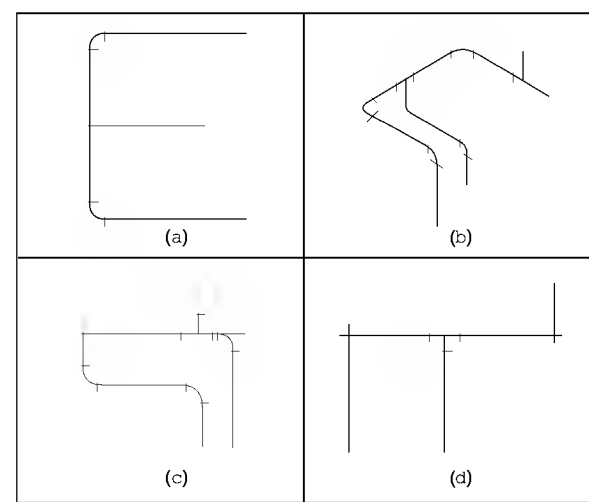


Fig. 7. Split screen viewing of designed parts. (a), Top view; (b), isometric view; (c), front view; (d), right-side view.

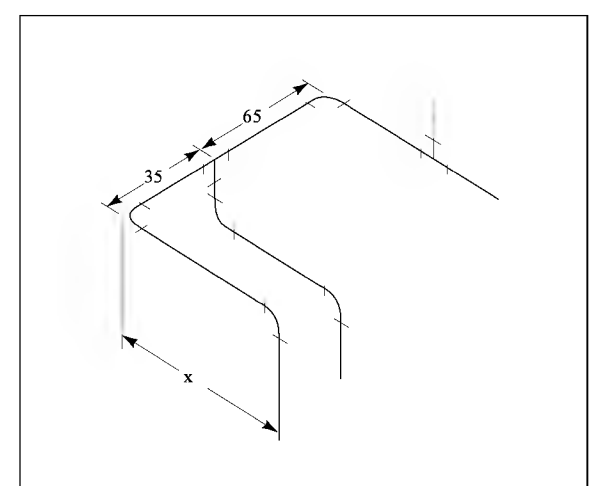


Fig. 8. Three-dimensional solid model.

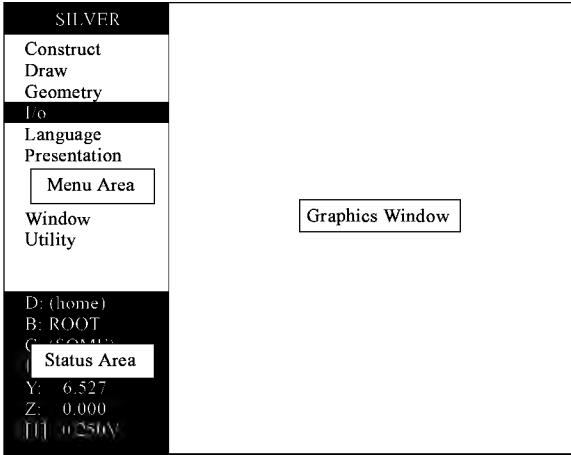


Fig. 9. Computer-aided drafting.

Other Capabilities

The creation and analysis of process flow sheets has become much easier because of the availability of automated systems to draw and revise them. The goal of the use of the flow sheet as the input for process simulation and for process control is likely to be achieved reasonably soon. The use of interactive graphic displays for process monitoring and control is pervasive today.

It is likely that there will always be a distinction between the way CAD CAM is used in mechanical design and the way it is used in the chemical process industry. Most of the computations required in mechanical design involve systems of linear or linearizable equations, usually describing forces and positions. The calculations required to model molecular motion or to describe the sequence of unit operations in a process flow sheet are often highly nonlinear and involve systems of mixed forms of equations. Since the natures of the computational problems are quite different, it is most likely that graphic techniques will continue to be used more to display results than to create them.

Computer graphics will continue to provide an effective tool for drawing chemical structures. Great improvement in software packages can also be expected. As illustrated in Figure 10, SilverScreen can list the needed modules (benzene, ring, bond, CH₃, and N) at the bottom of the screen, and select for copying, rotate if necessary, and then place them at desired locations, all interactively.

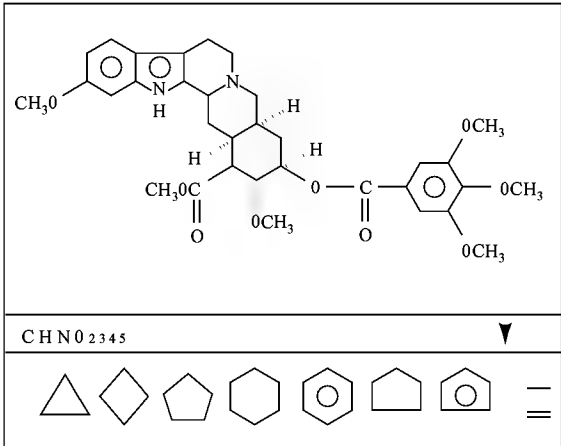


Fig. 10. Drawing chemical structures.

The use of color graphics is also an effective means for displaying chemical structures. This method is far better than typesetting the three-dimensional architecture of complex multimolecule assembly (112). For developing in-house CAD software programs, the three-dimensional, solid-modeling capabilities of SilverScreen can also be utilized either in monochrome or color for construction of such structures (113).

CAD CAM capabilities have had a subtle, but significant impact on chemical technology. Molecular modeling (qv) capabilities have led to the ability to describe chemical behavior and are showing promise in the design of new molecules. Pattern recognition techniques are being used to automate the analysis of chromatograms and spectral information.

It may be said that chemistry is about shapes, how things look, how things fit together, and how shapes change (114). For exploring noninteger dimensionality, fractals have emerged as a mathematical language. Unquestionably, CAD CAM applications will be further expanded in already established areas and will be intensively explored in many new fronts.

BIBLIOGRAPHY

1. Y. C. Pao, *Elements of Computer-Aided Design and Manufacturing*, John Wiley & Sons, Inc., New York, 1984.
2. M. F. Hordeski, *CAD/CAM Techniques*, Reston Publishing, Reston, Va., 1986.
3. M. P. Groover and E. W. Zimmers, Jr., *CAD/CAM Computer-Aided Design and Manufacturing*, Prentice-Hall, Englewood Cliffs, N.J., 1984.
4. R. V. Hughson, Supplement, *Chem. Eng. Prog.* **86**(10), 2-7 (1990).
5. E. M. Rosen, *Computer Applications to Chemical Engineering*, in R. G. Squires and G. V. Reklaitis, eds., *ACS Symposium Series*, vol. 124, Washington, D.C., 1980.
6. G. V. Reklaitis, R. S. H. Mah, and T. F. Edgar, *Eng. Educ.* **73**, 147-151 (1983).
7. T. F. Edgar, *CACHE News* (30), 32-34 (1990).
8. D. J. Wilde, *Globally Optimal Design*, Wiley-Interscience, New York, 1978.
9. C. J. Burt, *CACHE News*, (29), 13-16 (1989).
10. S. E. Zitney, R. D. LaRache, and R. A. Eades, *CACHE News* (31), 19-24 (1990).
11. B. Camahan, C. A. Reina, and E. J. Rennell, *MicroMENTOR System*, The University of Michigan, 1989.

12. J. S. Quarterman, *The Matrix: Computer Networks and Conferencing Systems*, Digital Press, Bedford, Mass., 1989.
13. Y. C. Pao and L. C. Chang, *Comput. Mech. Eng.* **3**(5), 65–72 (1985).
14. T. J. Schriber, *Simulation Using GPSS*, John Wiley & Sons, Inc., New York, 1974.
15. J. L. Humphrey, R. A. Koort, and A. F. Seibert, *CACHE News*, (32), 21–29 (1991).
16. S. Venkataraman, W. K. Chan, and J. F. Boston, *Chem. Eng. Prog.* **86**(8), 45–64 (1990).
17. P. R. Rony, *CACHE News* (32), (1991).
18. Y. C. Pao, *A First Course in Finite Element Analysis*, Allyn and Bacon, Boston, 1986.
19. Y. Cohen, W. T. Tsai, and S. Chetty, *CACHE News* (30), 18–26 (1990).
20. K. E. Gubbins and A. Z. Panagiotopoulos, *Chem. Eng. Prog.* **85**(10), 17–27 (1989).
21. D. M. Himmelblau, *Comput. Chem. Eng.* **5**(4), 187–195 (1981).
22. J. M. Calo and R. P. Andres, *Comput. Chem. Eng.* **5**(4), 197–209 (1981).
23. C. A. Eckert, T. T. Oberle, M. L. Gilbert, and R. S. Tomaskovic, *Comput. Chem. Eng.* **5**(4), 211–213 (1981).
24. M. Shacham and M. B. Cutlip, *Comput. Chem. Eng.* **6**(3), 209–214 (1982).
25. Ref. 24, pp. 215–218.
26. D. Gries, *Compiler Construction for Digital Computers*, John Wiley & Sons, Inc., New York, 1971.
27. N. C. Cohen, P. Collin, and G. Lemoine, *Tetrahedron* **37**, 1711 (1981).
28. N. Carpenter, *Comput. Chem. Eng.* **1**, 25–28 (1976).
29. J. Gioslowski and A. M. Turner, *Tetrahedron* **40**, 2161 (1984).
30. J. Gioslowski and A. M. Turner, *Comput. Chem. Eng.* **9**(4), 247–257 (1985).
31. W. P. Wipke, H. Braun, G. Smith, F. Choplin, and W. Sieber, *Computer Aided Organic Synthesis*, in W. P. Wipke and W. J. Howe, eds., *ACS Symposium Series* **61**, 97 (1977).
32. T. M. Dyott, A. J. Stuper, and G. S. Zander, *J. Chem. Inf. Comput. Sci.* **20**, 28 (1980).
33. J. Cohen, *A Compiler Generator*, Brandeis University, 1972.
34. N. A. B. Gray, J. G. Nourse, C. W. Crandell, D. H. Smith, and C. Djerassi, *Org. Magn. Reson.* **15**, 375 (1981).
35. T. Forester, B. Atzenberger, and C. W. van der Lieth, *Comput. Chem. Eng.* **12**(3), 267–283 (1988).
36. W. Linert, P. Margl, and E. Nusterer, *Comput. Chem. Eng.* **15**(1), 1–10 (1991).
37. E. G. Smith and P. A. Baker, *The Wiswesser Line-Formula Chemical Notation (WLN)*, 3rd ed., CIMI, N.J., 1975.
38. G. M. Come, C. Muller, P. Y. Cunin, and M. Griffiths, *Comput. Chem. Eng.* **8**(4), 233–237 (1984).
39. J. L. Bentley, L. W. Jelinski, and B. W. Kernighan, *Comput. Chem. Eng.* **11**(4), 281–299, (1987).
40. R. T. Haas and K. C. O’Kane, *Comput. Chem. Eng.* **11**(4), 251–271 (1987).
41. A. C. Norris and A. L. Oakley, *Comput. Chem. Eng.* **12**(3), 253–255 (1988).
42. R. C. Haines, *Comput. Chem. Eng.* **13**(2), 129–139 (1989).
43. L. Katz and C. Levinthal, *Ann. Rev. Biophys. Bioeng.* **1**, 465 (1972).
44. R. J. Feldmann, D. H. Bing, B. C. Furie, and B. Furie, *Proc. Nat. Acad. Sci.* **75**, 5409 (1978).
45. K. Knowlton and L. Cherry, *Comput. Chem. Eng.* **1**, 161 (1976).
46. C. K. Johnson, ORNL-5138, Oak Ridge National Laboratory, Oak Ridge, Tenn., 1976.
47. R. J. Feldmann, *Ann. Rev. Biophys. Bioeng.* **5**, 477 (1976).
48. D. D. Clark and S. M. Schuster, *Comput. Chem. Eng.* **4**, 83–85 (1976).
49. D. D. Clark and S. M. Schuster, *Comput. Chem. Eng.* **4**, 75–78 (1980).
50. D. D. Clark and S. M. Schuster, *Comput. Chem. Eng.* **4**, 79–82 (1980*b*).
51. J. Bauer and W. Schubert, *Comput. Chem. Eng.* **7**(2), 61–65 (1983).
52. M. Marsili, P. Floersheim, and A. S. Dreiding, *Comput. Chem. Eng.* **7**(4), 175–181 (1983).
53. M. L. Connolly and A. J. Olson, *Comput. Chem. Eng.* **9**(1), 1–6 (1985).
54. L. Watt and J. Kao, *Comput. Chem. Eng.* **9**(4) 269–277 (1985).
55. M. S. Newman, *Rec. Chem. Prog.* **13**, 111 (1952).
56. M. S. Newman, *J. Chem. Educ.* **32**, 344 (1955).
57. C. K. Johnson, ORNL-3794, Oak Ridge National Laboratory, Oak Ridge, Tenn., 1965.
58. G. M. Smith and P. Gund, *J. Chem. Inf. Comput. Sci.* **18**, 207 (1978).
59. D. R. Henry, *Comput. Chem. Eng.* **7**(3), 119–135 (1983).
60. M. Bishop and M. Csontos, *Comput. Chem. Eng.* **13**(1), 59–60 (1989).
61. F. C. Bernstein and co-workers, *J. Mol. Biol.* **112**, 535 (1977).
62. G. P. Zientara and J. A. Nagy, *Comput. Chem. Eng.* **7**(2), 67–74 (1983).
63. R. Rowlett, *Comput. Chem. Eng.* **9**(4), 301–315 (1985).
64. B. Clarke, *Comput. Chem. Eng.* **12**(1), 65–82 (1988).
65. F.-X. Bon and R. van Rapenbusch, *Comput. Chem. Eng.* **13**(4), 387–394 (1989).
66. N. L. Max, D. Malhotra, and A. Hopfinger, *Comput. Chem. Eng.* **5**, 19–27 (1981).
67. B. L. Bush, *Comput. Chem. Eng.* **8**(1), 1–11 (1984).
68. T. J. O’Donnell and A. J. Olson, *Comput. Graphics* **15**, 133 (1981).
69. D. C. Feller, E. F. Delmore, and S. S. Zimmerman, *Comput. Chem. Eng.* **13**(14), 337–341 (1989).
70. A. Kuwae, S. Obata, and K. Hanai, *Comput. Chem. Eng.* **9**(2), 109–113 (1985).
71. T. F. Edgar, R. Heeb, and J. O. Hougen, *Comput. Chem. Eng.* **5**(4), 225 (1981).
72. K. Mizutani, M. Toyama, K. Taguchi, and S. Mastumoto, *Comput. Chem.* **9**(4), 259–267 (1985).
73. R. E. Carnhart and D. H. Smith, *Comput. Chem. Eng.* **1**(2), 79–84 (1976).
74. M. E. Munk and B. D. Christie, *Anal. Chem. Acta* **216**, 57–68 (1989).
75. B. D. Christie and M. E. Munk, *J. Am. Chem. Soc.* **113**, 3750–3757 (1991).
76. K. K. Agarwal, D. L. Larsen, and H. L. Geletner, *Comput. Chem. Eng.* **2**, 75–84 (1978).
77. E. J. Corey and W. T. Wipke, *Science* **166**, 178 (1969).
78. J. B. Hendrickson, *Topics Curr. Chem.* **62**, 49 (1976).
79. I. Ugi and co-workers, *Angew. Chem.* **91**, 99 (1979).
80. T. Suzuki and M. Ishida, *Comput. Chem. Eng.* **14**(3), 211–217 (1990).
81. R. C. Reid, J. M. Prausnitz, and T. K. Sherwood, *The Properties of Gases and Liquids*, McGraw–Hill, New York, 1977.
82. J. T. Adams and E. M. T. So, *Comput. Chem. Eng.* **9**(3), 269–284 (1985).
83. K. G. Joback and G. Stephanopoulos, *Foundations of Computer-Aided Process Design*, J. J. Sirola, I. E. Grossmann, and G. Stephanopoulos, eds., Elsevier, New York, 363–387, 1990.
84. M. Ishida and D. Zheng, *Comput. Chem. Eng.* **10**(6), 525–532 (1986).
85. P. Gans, *Comput. Chem. Eng.* **1**, 291–293 (1977).

86. P. Borrell, *Comput. Chem. Eng.* **4**, 131–135 (1980).
87. M. Albin and H. B. Gray, *Comput. Chem. Eng.* **13**(3), 173–177 (1989).
88. V. N. Schrodtt and A. M. Saunders, *Comput. Chem. Engr.* **5**(4), 299–305 (1981).
89. E. Fantini, L. Tognotti, and A. Tonazzini, *Comput. Chem. Eng.* **14**(11), 1201–1211 (1990).
90. R. S. Brodkey, *CACHE News*, (28), 19–21 (1989).
91. K. Yamamoto, H. Ozoe, P. Chao, and S. W. Churchill, *Comput. Chem. Eng.* **6**(2), 161–167 (1982).
92. W. B. Cheng and R. S. H. Mah, *Comput. Chem. Eng.* **2**, 133–142 (1978).
93. B. Curtis, M. K. Sood, and G. V. Reklaitis, *Comput. Chem. Eng.* **5**(4), 277–287 (1981).
94. M. K. Sood, T. Wang, and G. V. Reklaitis, *Comput. Chem. Eng.* **5**(4), 289–298 (1981).
95. S. P. Singh and B. Carnahan, *Comput. Chem. Eng.* **5**(4), 261–276 (1981).
96. P. E. Preece, S. A. al Bader, J. L. Davila Pernia, J. A. C. Evans, and M. K. Giller, *Comput. Chem. Eng.* **11**(3), 279–289 (1987).
97. A. B. Guillory and F. C. Knopf, *Comput. Chem. Eng.* **11**(5), 527–525 (1987).
98. S. J. Sawicki, R. D. Young, and S. E. Sund, *Comput. Chem. Eng.* **10**(3), 297–301 (1986).
99. L. B. Evans, *Chem. Engr. Prog.* **86**(10), 42–44 (1990).
100. T. F. Edgar, R. Heeb, and J. O. Hougen, *Comput. Chem. Eng.* **5**(4), 225 (1981).
101. W. G. Etzkorn and Y. Arkun, *Comput. Chem. Eng.* **5**(4), 233 (1981).
102. D. Buatala, K. Y. Choi, and M. K. H. Fan, *Comput. Chem. Eng.* **12**(11), 1115–1127 (1988).
103. B. Linnhoff and J. R. Flower, *AIChE J.* **24**, 633 (1978).
104. A. K. Saboo, M. Morari and R. D. Colberg, *Comput. Chem. Eng.* **10**(6), 577–589 (1986).
105. Ref. 104, pp. 591–599.
106. G. R. Strickler, *Chem. Eng. Prog.* **82**(12), 50–56 (1991).
107. J. Calandranis, G. Stephanoloulos, and S. Nunokawa, *Chem. Eng. Prog.* **86**(1), 60–68 (1990).
108. R. V. Hughson, *Chem. Eng. Prog.* **87**(11), 68–71 (1991).
109. Y. C. Pao, *CoED J.* **5**, 20–24 (1985).
110. Y. C. Pao, *Proc. Int. Computers in Eng. Conf. Exhib.* **3**, 15–20 (1986).
111. F. M. Julian, *Chem. Eng. Prog.* **85**(10), 33–40 (1989).
112. R. Dagani, *Chem. Eng. News* **69**(21), 24–30, (1971).
113. Y. C. Pao and M. Foltz, *Engineering Drafting and Solid Modeling with Silver Screen*, CRC Press, Inc., Boca Raton, Fla., in press.
114. S. Borman, *Chem. Eng. News* **69**(16), 28–35 (1991).

Y. C. Pao
L. D. Clements
University of Nebraska, Lincoln

COMPUTER-AIDED ENGINEERING (CAE)

Engineering may be defined as a creative activity whereby a person uses the laws of nature and the knowledge of science for the benefit of humanity, while taking economic considerations into account (1). The laws of science, especially those of physics and chemistry, are generally stated as mathematical equations. Thus, engineering consists of applying those mathematically stated laws to the design and analysis of systems for useful purposes. Specifically, an engineer builds a mathematical model of an engineering system, such as a reactor or a chemical processing plant, in terms of the laws of physics and chemistry; then determines the parameters for the model; and finally, manipulates the mathematical model to arrive at conclusions about that system, such as its performance or its design.

A significant part of this work involves making engineering calculations, in other words, finding numerical solutions to a set of mathematical equations. The equations used may be linear or nonlinear and may or may not involve the derivatives and integrals of variables. The variables themselves may be discrete or continuous. From the time when computers were first available, engineers have used them to perform such numerical computations. Some unique problems arise because computers represent numbers electronically in binary form as they perform mathematical operations on the numbers stored in their memory. For example, on most computers $(\frac{1}{3}) \times 3$ will not produce 1 as a result, but will give 0.9999999; the number of 9s depends on the numerical accuracy of the electronic circuits of the computer. Such numerical errors are frequently magnified when one uses a computer to solve systems of nonlinear and differential equations. The discipline of numerical analysis has arisen to deal with such issues. It is often necessary to perform numerical analysis when developing computer programs; sometimes it must even be applied when using them to solve complex engineering problems.

Whenever a computer is used for engineering computations, several layers of programming are involved. Figure 1 represents those layers and their relationships with each other. Computer hardware cannot function for any practical purpose without an operating system that manages the various components of the hardware. By and large, modeling systems are written in higher level languages, which are then compiled (or translated) into a computer's machine language before the operating system presents them to the machine. *FORTRAN* is the most popular language for writing engineering programs because of its ability to represent mathematical expressions easily. *BASIC* and *C* are two other languages that engineers employ increasingly, *BASIC* because it is more readily available for use on personal computers and microprocessors and *C* because of its general used by UNIX, an operating system developed and promoted by AT&T. Furthermore, *C* has more direct access to the computer's hardware.

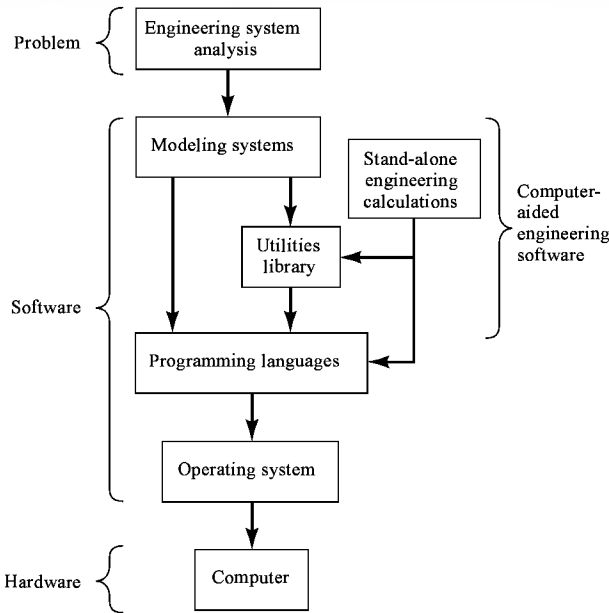


Fig. 1. Layers of programming with engineering computation software.

Figure 2 depicts the entire process of computer-aided engineering (CAE). A part of this process is the software for computer-aided engineering, which, as shown in Figure 1, consists of modeling systems, stand-alone programs, and libraries of utilities for engineering calculations.

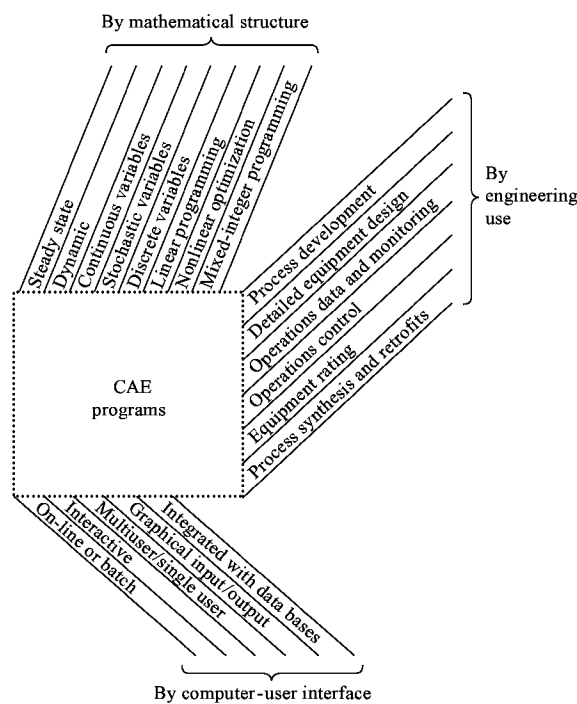


Fig. 2. Ways of categorizing CAE programs.

Engineers and technical professionals can write their own programs in higher level languages for simple calculations, making use of utility libraries, such as a general-purpose routine for finding the roots of a quadratic equation. However, for analyzing complex engineering systems, professionals use flexible, general-purpose modeling systems. For process simulation, for example, there are systems available that are themselves computer programs written in one of the higher level languages, usually *FORTRAN* or *C*. The technical professionals who are the end users model their engineering problems by using the modeling system already developed for a particular problem. Certain special-purpose simulation languages such as *SLAM*, *GPSS*, and *SIMSCRIPT*, also are used to model systems, but their application covers only a limited domain of problems.

A modeling system is a computer program that allows numerical representation of the mathematical model of an engineering system. A number of modeling systems have been developed that produce different types of models and that employ different approaches to modeling. Modeling systems can be classified according to the structures and types of the mathematical models they are designed to analyze. Thus there are modeling systems for steady-state and time-varying–dynamic systems, for linear and nonlinear equations, for linear programming and for optimizing with nonlinear constraints and nonlinear objective functions, for modeling stochastic and deterministic processes, for modeling with continuous and with discrete and integer variables, etc.

Another way to categorize engineering computation programs is to consider their engineering function. Thus there are programs for designing new equipment and for rating existing equipment, for designing manufacturing plants and for managing and operating existing plants. Often, the same modeling systems can be used for both design and operation. Figure 2 depicts these different ways of characterizing CAE modeling systems.

The availability of microprocessors and desktop computers that serve as engineering workstations has recently dramatically increased the use of computers in engineering (2). In addition, general-purpose generic software (spreadsheets, data base systems, and interactive graphics packages) is used increasingly for storing and analyzing engineering data.

When computers were first used in engineering, they performed numerical computations and processed information by manipulating strings of characters. Rapid significant advances in the science of programming and in understanding the nature of information have opened up the use of computers for working with large volumes of nonnumeric information, both symbolic and pictorial. Whereas computers once solved a set of algebraic equations numerically, some newer software packages can manipulate algebraic equations symbolically to obtain an analytical solution. Differentiation and integration can be carried out for many problems analytically, using computer software. Process simulation systems previously created process flow diagrams by using tables, which listed the streams that flowed in and out of specific process equipment. It is now becoming common instead to construct the flow diagrams on the computer screen graphically and have that drawing provide the input to the process simulation system.

Process Simulation

At the simplest level, process simulation involves making material and energy balances for a process flow sheet. Generally, it involves much more detailed modeling of any type of process for which there is a continuous flow of materials and energy from one processing unit to the next (3–5). Simulators have been used to model processes in chemical and petrochemical industries, petroleum refining, oil and gas processing, synthetic fuels, power generation, metals and minerals, pulp and paper, food, pharmaceuticals, and biotechnology.

Flow-sheet models are used at all stages in the life cycle of a process plant during process development, for process design and retrofits, and for plant operations. Input to the model consists of information normally contained in the process flow sheet. Output from the model is a complete representation of the performance of the plant, including the composition, flow, and properties of all intermediate and product streams and the performance of the process units.

During process development, a model can be developed as soon as a conceptual flow sheet has been formulated. This model can be updated as more information about the process is obtained. Even at an early stage in the project, the model can be used to assess the preliminary economics of the process and the effect of technological changes on these economics. The model can aid in interpreting pilot-plant data and allows the study of many process alternatives.

During process design, once the decision has been made to build a new plant or to modernize an existing plant, simulation models can be used to study trade-offs, to investigate off-design operation, and to analyze the flexibility of the plant to handle a range of feedstocks. Simulation studies during process design can avoid costly mistakes before committing to plant hardware. Process engineers can use a simulation model to optimize the design of the process by making a series of case studies to ensure that the plant will work properly under a wide range of operating conditions.

For an existing plant, a model can serve as a powerful tool for plant engineers to improve plant operations, to enhance yield and throughput, and to reduce energy use. The model can be used to determine changes in operating and safety conditions needed to accommodate changes in feedstocks, changes in product requirements, and changes in environmental conditions. The model can guide plant operations to reduce cost and improve productivity. Finally, the model can be used to study possible plant modifications for removing bottlenecks or for revamping the plant to incorporate technology advances such as an improved catalyst, a new solvent, or a new process unit.

Using simulation offers several advantages (6). Simulation makes it possible to investigate and experiment with the complex internal interactions of the system being studied whether a single process, an overall plant, or a complete company. It allows investigation of the sensitivity of a system to small

changes in internal parameters or environmental conditions and thus helps determine the accuracy to which these factors must be known. Simulation also allows the testing of the model system in regimes of operation that would be too costly, too dangerous, too time-consuming, or beyond the operating ranges of any one particular physical system, ie, systems sizes, construction materials, etc. Because they obviously cause no disturbances in the real system being studied, investigations at the limits of operation of the real system can be made with impunity, without danger to the system or its human operators.

For certain types of stochastic or random-variable problems, the sequence of events may be of particular importance. Statistical information about expected values or moments obtained from plant experimental data alone may not be sufficient to describe the process completely. In these cases, computer simulations with known statistical inputs may be the only satisfactory way of providing the necessary information. These problems are more likely to arise with discrete manufacturing systems or solids-handling systems rather than the continuous fluid-flow systems usually encountered in chemical engineering studies. However, there are numerous situations for such stochastic events or data in process industries (7–10).

The steps involved in developing a flow-sheet model are as follows.

- (1) Define the process flow sheet to be modeled and the purpose of the model.
- (2) Select the units of measurement for input data and output reports.
- (3) Specify the chemical components present in the streams of the flow sheet.
- (4) Specify the methods and models to be used for calculating physical properties.
- (5) Break the process flow sheet into unit operations and choose an appropriate model for each unit.
- (6) Specify the performance of each unit operation to represent the design and operating conditions of the process.
- (7) Impose design specifications.
- (8) Set up sensitivity analyses or case studies.

Steps 1, 2, and 3 involve engineers' definitions for the problem that is to be tackled. Practically all the simulators have special menus or special input forms displayed on screens for entering appropriate information into the computer program.

Step 4 deals with physical and chemical properties of compounds and mixtures. Accurate physical and chemical properties are essential to achieve accurate simulation results. Most simulators have a method of maintaining tables of these properties as well as computer routines for calculations for the properties by different methods. At times these features of simulators make them suitable or not suitable for a particular problem. The various simulators differ in the number of compounds in the data base; number of methods for estimating unknown properties; petroleum fractions characterized; electrolyte properties handled; biochemical materials present; ability to handle polymers and other complex materials; and the solids, metals, and alloys handled.

Step 5 deals with the process flow diagram. The calculations procedure that the early simulators used, still the most common in industrially used simulators, consists of making calculations around each unit operation of a process in a sequence of unit operations through which the main material flows. These simulators are known as block sequential or sequential modular simulators. The topology of the process (including all recycles and purges) must be given to the simulator, each of which has an equivalent way of entering inputs and outputs of each unit operation or source and destination of each stream. There is a trend toward enabling an engineer to draw the flow diagram on a computer screen and have the computer program pick up the needed information about the topology of the process directly from the drawing.

Earliest simulators determined the sequence of the calculations for the various unit operations from the explicit input from the engineer or from the sequence in which the topological information about unit operations was entered. All simulators today analyze the topology automatically and determine the sequence. Most of them, however, allow the user to alter this sequence through various techniques of creating hypothetical calculation units of convergence, recycle, and control blocks or units.

The topology for calculations is similar to the topology of the process flow sheet, but not exactly the same (4). The simulators require calculations flow diagrams. On a process flow diagram, combining two streams may be shown as indicated in Figure 3, without any unit operation or processing equipment. For a simulator, this situation must be represented as shown in Figure 4 where the two streams are mixed in a hypothetical mixer.

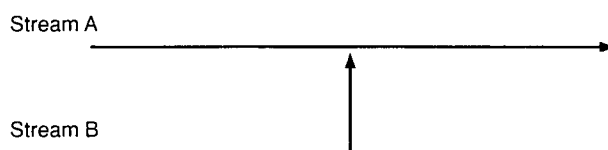


Fig. 3. Topology of two combined streams on a process flow sheet.

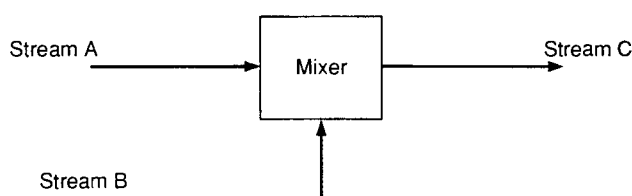


Fig. 4. Topology of two combined streams in a simulation.

Similarly a hypothetical splitter must be defined when a stream is split. Furthermore, control units are identified in calculations flow diagrams when temperatures, pressures, and flows of streams are controlled on the basis of variables in other parts of flow sheets. Sometimes variables such as product specifications and reactor yields can be represented as hypothetical control units.

Most flow sheets have one or more recycles, and trial-and-error becomes necessary for the calculation of material and energy balances. The calculations in a block sequential simulator are repeated in this trial-and-error process. In the language of numerical analysis, this is known as convergence of the calculations. There are mathematical techniques for speeding up this trial-and-error process, and special hypothetical calculation units called convergence, or recycle, units are used in calculation flow diagrams that invoke special calculation routines.

There is a great difference between various simulators (5) in terms of how easily and how well the hypothetical calculation units can be incorporated in the simulation. The trial-and-error calculations, which are called iterative calculations, do not always converge for every flow sheet being simulated. Test problems can be devised to be tried with various simulators to see if the simulator will give a converged solution (11). Different simulators could take different numbers of iterations to converge and take different amounts of computer time on the same computer.

Steps 6 and 7 are involved with inputting additional specifications about the process being simulated. It is necessary to give an adequate number of specifications for each unit operation, for each calculation unit, and for the overall process flow so that all the degrees of freedom are taken away and a unique solution can be obtained from the simulator. On the other hand, if more than the necessary number of specifications are given, the problem becomes overconstrained for the simulator and no solution can exist.

In a large number of processes, there are unit operations related to vapor-liquid separations: distillation, absorption, extraction, stripping, flashing, and separation of liquid and vapor stream arising from changes in temperatures and pressures. Calculations for these unit operations necessitate trial and

error, and all simulators have models for several such operations. The degree of details and flexibility of specifications for these units as well as the calculation procedures used for these units differ significantly in various simulators.

Distillation calculations can be sensitive indicators of how well a simulator fits a problem and how well it performs. Models for simulation of petroleum crude units and vacuum distillation of heavy petroleum fractions are necessary for applications in the oil and gas industry. Some distillation calculations will not converge for systems of chemicals that are close to boiling or form azeotropes. The number of iterations and the computer time for convergence of distillation calculations (and whether they converge at all or not) also is greatly influenced by the physical properties and the methods of estimating them. The greater the nonlinearity and nonideality of the properties of mixtures of compounds involved, the greater the difficulty of convergence. The specification of the distillation operations, the size of model used, and starting guesses for the calculations also influence the convergence; this influence can be different for different simulators.

Equations-Oriented Simulators. In contrast to the sequential-modular simulators that handle the calculations of each unit operation as an input–output module, the equations-oriented simulators treat all the material and energy balance equations that arise in all the unit operations of the process flow sheet as one set of simultaneous equations. In some cases, the physical properties estimation equations also are included as additional equations in this set of simultaneous equations.

The essential differences between sequential-modular and equation-oriented simulators are in the structure of the computer programs (5) and in the computer time that is required in getting the solution to a problem. In sequential-modular simulators, at the top level, the executive program accepts input data, determines the flow-sheet topology, and derives and controls the calculation sequence for the unit operations in the flow sheet. The executive then passes control to the unit operations level for the execution of each module. Here, specialized procedures for the unit operations library calculate mass and energy balances for a particular unit. Finally, the executive and the unit operations level make frequent calls to the physical properties library level for the routine tasks, enthalpy calculations, and calculations of phase equilibria and other stream properties. The bottom layer is usually transparent to the user, although it may take 60 to 80% of the calculation efforts.

In the equation-oriented approach, the executive organizes the equations and controls a general-purpose equation solver. The equations for material and energy balances may be grouped separately from those for the calculation of physical properties or phase equilibria, or as in the design of some simulators, the distinction between these groups of equations may disappear completely.

The computer effort required for convergence depends on the number and complexity of the recycles in the flowsheet, the nonlinearities in the physical properties, and the nonlinearities in the calculation of phase or chemical equilibria. In sequential-modular simulators these calculations are converged one at a time, sequentially, and in a nested manner. In equation-oriented simulators they are converged as a group and, in the case of complex flow sheets involving nonideal mixtures, there could be significant reduction in computer effort.

Historically, sequential-modular simulators were developed first. They were also developed primarily in industry. They continue to be widely used. In terms of unit operations, each module can be made as simple or complex as needed. New modules can be added as needed. Equation-oriented simulators, on the other hand, are able to handle arbitrary specifications and limitations for the entire process flow sheet more flexibly and conveniently than sequential-modular simulators, and process optimization can also be carried out with less computer effort.

Convergence Methods. In the simplest terms, convergence means finding a solution to a trial-and-error calculation. The simplest procedure is direct substitution of the recalculated value in the trial-and-error calculation. There are several mathematical and algorithmic techniques (studied in the field of numerical analysis) of speeding up this convergence process by providing a better guess for each succeeding iteration than direct substitution. Newton's method is most commonly the first one attempted, because it is simple and, in many situations, effective. It consists of linearizing the equations based on derivatives at the current point of iteration and then solving the linearized simultaneous equations. The solution provides the guess for the next iteration. For highly nonlinear situations and when the functions have many inflections, Newton's method could run into problems of cycling, and Wegstein's and other methods are used instead.

Numerical techniques for improving convergence can be applied to both equation-oriented and sequential-modular simulations. Simultaneous equation solution modules of equation-oriented simulators bring about simultaneous convergence of many variables using these methods. In sequential-modular simulators, special convergence blocks (hypothetical operations) are introduced to enable simultaneous convergence of several variables, thereby speeding up calculation of nested recycles enough to be comparable with the equation-oriented approach.

The computer effort required to get a solution to a simulation problem is important because, in the cases of optimization of design and dynamic simulation for control, many simulator runs must be made. At times the models of process units are simplified and often linearized to speed up the convergence.

Steady-State Simulation Programs. The earliest process simulators were developed by petroleum (Exxon and Chevron) and chemical (Union Carbide, Monsanto, and Imperial Chemical) companies and by engineering construction contractors (Kellogg) (12,13). Many of these simulators contained proprietary information about the properties of some of their materials and were customized to handle proprietary models of some of their processes. Monsanto, however, gave their *FLOWTRAN* program to universities for use in teaching. Research in universities produced *PACER* at Dartmouth, *ASPEN* at MIT, and *CHESS* at Houston. Cambridge University and Imperial College in the UK, and Carnegie Mellon in the United States have championed the equation-oriented approaches. A number of independent software companies now develop and support high quality simulation software, and the trend has been for companies that previously developed their own simulators to rely on these independently produced simulation packages. Table 1 lists these companies and their software.

Table 1. Comparison of Simulator^a Capabilities

Characteristics	Simulator							
	ASPEN PLUS	Chemcad II	Design-II	Hysim	Pro-II	Quasilin	Span	SPEEDUP
company	Aspen Technology	Coade Chem Stations	Chem-Share Corp.	Hyprotech Inc.	Simulation Sciences Inc.	Lynxvale Ltd.	Kesler Eng. Inc.	Aspen Technology UK Ltd.
hardware ^a	1–5	1	1–4	1	1–5	2–4	1,3	2–5
<i>Physical property capabilities^b</i>								
methods and data available	1,4,5,10–13	1,4,5,7,10–13	1,3–5,7,10–13	1,4,5,7,10–12	1,10–13	7,9	10–12	6–8
handle electrolytes?	Y	Y	Y	N	Y	N	N	Y
handle petrol. fractions?	Y	Y	Y	Y	Y		Y	Y
add own estim. methods?	Y	Y		Y	N		N	Y
data regression?	Y	Y	Y	Y	Y	N	Y	Y
<i>Unit process modeling</i>								
reactor models available ^c	A,B,C	A,B,C	A,B	A,B	A,B,C	A	A	A,B
incorporate user modules?	Y	Y	Y	Y	Y	Y	N	Y
handle batch reactor (R) or batch distillation (D)?	Y	Y	Y	N	D only	Y	N	Y
handle streams with solids?	Y	Y	N	Y	Y	N	N	Y
<i>Input/output features</i>								
accept graphical input as	Y	Y	Y	Y	Y		N	Y

flow diagram?								
produce PFD ^d as output?	Y	Y	Y	N	Y	N	N	Y
plot curves of simulation results and sensitivity studies?	Y	Y	Y	Y	Y	Y	N	Y
company	Aspen Tech-nol-ogy	Coade Chem Stations	Chem-Shar e Corp.	Hyp-ro-te ch Inc.	Simula-tio n Sci-ences Inc.	Lynxvale Ltd.	Kesler Eng. Inc.	Aspen Techn-ology UK Ltd.
			System modeling features					
optimizer available?	Y	N	Y	Y	Y	Y	Y	Y
dynamic stimulator available?	N	N	Y	Y	Y	Y	N	Y
algorithm used ^e	SMD	SMD	SM	HS	SM	EO	SM	EO
			Other features					
has project design database?	Y	N	Y	Y	Y	Y	Y	uY

^a Computer platforms: where 1 represents IBM PCs and compatibles, 386 or better; 2, Unix-based work stations (or equivalent) including HP Apollo, SUN, and IBM RS6000; 3, DEC VMS; 4, IBM mainframe, MVS or VM, or compatible; and 5, others including Cray, Data Gen., etc.

^b Physical property methods and data include 1, activity coefficient correlation; 2, ChemAdat; 3, ChemTran; 4, DIPPR; 5, Group Contribution; 6, PPDS; 7, Prausnitz; 8, AspenTech's PROPERTIES PLUS; 9, Thermopack; 10, UNIFAC; 11, UNIQUAC, Van Leer, Wilson, and NRTL; 12, Soave-Redlich-Kwong, Peng-Robinson Eq.'s of S; and 13, Chao-Seader, Grayson Streed plus one or more of Braun K-10, B-W-R, and other Eq.'s of S.

^c Reactor types modeled: A, stoichiometric conversion; B, equilibrium free-energy minimization, continuous stirred tank, and plug flow; C, reactive distillation. Some vendors have special models for special reactions; also, private company simulators usually have reactors of specific interest to their company.

^d PDF process flow diagram.

^e Algorithmic structure of simulator: EO, equation oriented; HS, hybrid system; SM, sequential modular; and SMD, sequential modular, suitable for design.

Many sophisticated simulators have been developed by chemical companies, engineering contractors, and refiners. It may be possible to arrange for their use, but they are not standard, commercially available products. Reference 5 offers an overview of their features and capabilities. Examples include CAPES, Chiyoda Corp. (Yokohama, Japan); CHEMASIM, BASF (Ludwigshafen, Germany); FLOWPACK, ICI Engineering (Cheshire, UK); GENESIS, BP International, Ltd. (London); IPES, Union Carbide Corp. (South Charleston, West Virginia); MPFPL, OPTISIM, Linde AG (Hoellriegelskreuth, Germany); and VTPLAN, Bayer AG (Leverkusen, Germany). There are also modules of independent programs available from which simulations can be synthesized, for example, the set of programs available from ChemEng Software and Services, Ltd. (Dorset, UK).

Physical Properties

A process-simulation program almost always contains a physical property service, because the quality of process design ultimately depends on the way in which the laws of physics and chemistry are applied to the problem. Accordingly, the quality of this service is an important consideration to the user of a flow-sheeting system.

The physical property service must perform a number of tasks, but the most useful of these are the following:

- (1) To supply estimates repetitively for a number of different physical properties while the simulation is in progress.
- (2) To provide the user with the values of properties of interest during the calculation and at simulation completion, for subsequent use in other calculations.
- (3) To allow the user to input special data for new components and transform them into the form required by the system during a simulation.
- (4) To supply the user with a means to estimate properties when little, except perhaps chemical structure, is known about a particular chemical compound; again the system must put these estimates into the form needed by the simulation during execution.

General Properties of Computerized Physical Property System. Flow-sheeting calculations tend to have voracious appetites for physical property estimations. To model a distillation column one may request estimates for chemical potential (or fugacity) and for enthalpies 10,000 or more times. Depending on the complexity of the property methods used, these calculations could represent 80% or more of the computer time required to do a simulation. The design of the physical property estimation system must therefore be done with extreme care.

Figure 5 summarizes the general features of what one may expect to find in a physical property system. First consider the central column of the diagram, which represents the basic set of facilities.

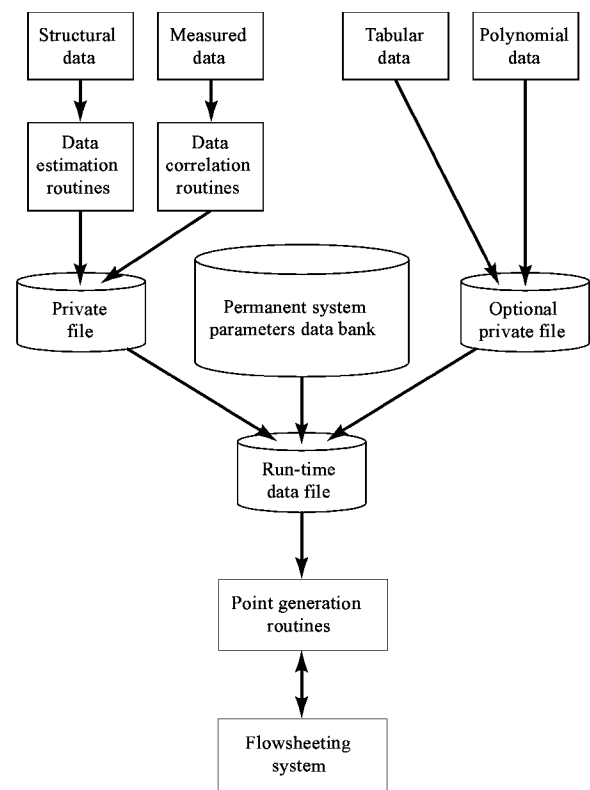


Fig. 5. General features of a physical property system.

The simulation models of the flow-sheets system must make frequent requests for properties at specific temperatures, pressures, and compositions. Computer-program calls for such data are usually made in a rigorously defined manner, which is independent of both the point data generation models and the particular components. These point generation routines provide the property values, using selected methods that base their calculations on a set of parameters for each component.

At the time of a computer run of a simulation, it is desirable to store only the parameters for those particular components involved in the simulation, and so this set of basic data is normally copied over from the permanent system parameters data bank into run-time data locations. In typical flow-sheets problems this involves the collection of particular parameters for 5 to 20 components from a data bank with extensive sets of data for up to several thousand components.

Use of such a data system is easy, provided that the permanent system data bank has entries for the components of interest. Thus it is often important to the user that the data bank is extensive and not restricted to a small class of compounds. However, the number of chemical species is enormous and expanding at a rapid rate. Accordingly, for wide application it is necessary that the physical property system embodies facilities for the use of user-supplied data.

This requirement of a physical property system is generally accommodated by enabling the user to create the equivalent of the permanent system data bank by explicitly entering data in the same format. Such private data banks can be used independently or in conjunction with the system data bank. Data supplied in this way normally require that the user has access to expertise in physical property data and possibly in computer use as well. In these circumstances the user may also wish to provide data in tabular or polynomial form for use by an appropriate set of interpolative point generation routines. This facility is shown at the top right of Figure 5.

In addition to these facilities for supply of data in an explicit form for direct use by the system, there also are options designed for the calculation of the parameters used by the system's point generation routines. Two obvious categories of this type can be identified and are included at the top left of Figure 5. The first of these applies to the correlation of raw data and is most commonly applied to the estimation of binary interaction parameters.

The second category differs from those discussed above in that it relates, in the main, to those situations for which no data or only characterizing data exist. In such cases, this small set of characterizing data or, in its absence, structure data are used to estimate a set of parameters of the type required by point generation routines. One notable specific example of this type of facility is the creation of data sets for petroleum boiling fractions from information on average boiling point and density.

All the foregoing facilities form part of the spectrum of options that, in addition to the permanent system data bank, enable the engineer to get the most out of a flow-sheets system. The following list shows the physical properties that are often required for process simulation. The methods of estimating these properties, when direct measurements are not available, are indicated in the references following the properties (also see THERMODYNAMIC PROPERTIES).

Thermodynamic

- Equations of state (14–18)
- Density (19,20)
- Compressibility
- Thermal expansion coefficients
- Heat capacity
- Enthalpy
- Heat of mixing
- Entropy
- Gibbs free energy
- Heats of formation
- Entropy of formation
- Heat of reaction
- Free energy of reaction
- K-values (vapor–liquid equilibrium)
- Activity coefficients (21–26)
- Fugacity coefficients
- Chemical potential

- Fugacity
- Transport
- Viscosity (27–34)
- Thermal conductivity (35–37)
- Diffusion coefficients (38–41)
- Surface tension

Process Equipment Design and Rating

There is a wide array of computer programs that have been developed for the sizing and detailed mechanical design of processing equipment such as distillation and other separation columns, heat exchangers, compressors, etc. Process simulation programs described previously define the material and energy flows through such pieces of equipment for a particular process. The next step in the design of any process is the sizing of the equipment.

Some process simulators have sizing calculations built into the unit operation models. For example, having determined the flows and internal refluxes in a distillation column, the user can then calculate the diameter of a bubble cap or a packed column. Generally, the sizing calculations do not interact with the material and energy balances and are invoked by the engineer through calling special routines after the material and energy balances have been completed.

For many pieces of equipment, such as heat exchangers and distillation columns, stand-alone programs are available that calculate material and energy balances around that piece of equipment, size the equipment, and calculate or rate its performance.

Process simulators stop generally at the process specifications for the equipment. For the detailed mechanical design of the equipment, such as heat exchangers and distillation columns, stand-alone programs are often used. They make process calculations, size the equipment, calculate thermal and mechanical stresses, design mechanical support of the parts of the equipment, design inlet and outlet nozzles, etc.

Some computer programs can both design and rate equipment, whereas some can only design or rate the performance of the equipment. For the design of equipment, the process conditions are given as input to the program, which then produces all the size parameters of the equipment as the output. In rating, on the other hand, the physical dimensions or the size of existing equipment are input into the program, which then predicts the performance of the equipment and the outlet conditions of the process streams from the equipment. Rating programs are useful for studying alternative uses of existing equipment for purposes other than what it might have originally been designed for, and for monitoring the internal condition of equipment in service, such as the fouling of a heat exchanger or proper operation of trays or packing in separation columns. They can also be used to identify material and energy losses arising from internal breakdowns of equipment such as tube ruptures.

For the pieces of equipment that are somewhat standard items, such as pressure-relief valves, heat exchangers, pumps, etc, there is a trend toward the manufacturer of the equipment supplying computer programs for equipment selection. For more complex process operations, and for more elaborate calculations for the simpler equipment, software may be available from a number of engineering software vendors. Engineering contractors and large manufacturing companies often have their own software that incorporates the design or the operations standards of the company.

Heat Exchangers. Because a heat exchanger is a nearly universal item of process equipment, there is a wide variety of fast and capable software for such calculations.

All heat exchanger programs carry out thermal calculations of heat transfer through the exchanger surface area, but they differ in details of these calculations. Some handle condensing and boiling fluids, others do not. Some calculate the heat-transfer coefficients; others require its value to be input by the engineer. Many programs rate specific types of heat exchangers. Most handle shell-and-tube exchangers, but once again they differ in details. For example, some consider the baffles on the shell side in the calculations for the heat-transfer coefficients, whereas others do not.

Some programs deal only with thermal design; others carry out mechanical design such as tube layout, baffle sizes, spacing, etc for the shell-and-tube exchangers. Once again, they differ in the degree of details in the mechanical design.

If the heat transfer required for cooling or heating a stream is large, a set of smaller heat exchangers may be more economical than a single large exchanger. In such a case, the problem of optimizing the total cost or the total surface area arises. Some programs determine the number of heat exchangers by optimization, whereas others do not.

In complex plants such as petroleum refineries, a number of process streams need cooling and heating. Several questions arise: Which stream should exchange heat with which stream? How should a single stream be routed through a series of heat exchangers? How much heat should be exchanged so as not to reduce the temperature driving force to the point of uneconomical use of the heat-transfer surface area? Some special programs handle these problems in various degrees. One method known as the analysis of "pinch" enables an engineer to design a heat-exchanger network in such a way that there is no crossover of temperatures and a reasonable driving force of temperature difference is maintained in each heat exchanger. This method is essentially a graphical method, but it is implemented using special computer systems and programs. Table 2 lists some heat-exchanger packages.

Table 2. Heat-Exchanger Packages

	HEXTR AN	ADVEN T	HEXNA N	HETRAN teams	Chemical c 6	HTC-ST X	PCI-HEX N	Energy analyst	SUPE R-TAR GET	Chemcal c 5
reference	Simulati on Sciences, Inc.	Aspen Technolo gy, Inc.	Kesler Engineeri ng, Inc.	B-Jac Internatio nal	Gulf Publishin g Co.	Heat Transfer Consultan ts, Inc.	PCI Consultan ts, Inc.	Thermal Analysis Systems Co.	Linnho ff March, Inc.	Gulf Publishin g Co.
location	Fullerto n, Calif.	Cambrid ge, Mass.	East Brunswick , N.J.	Midlothian , Va.	Houston, Tex.	Pleasant Hill, Calif.	Houston, Tex.	Elk Grove Village, Ill.	Leesbu rg, Va.	Houston, Tex.
computer on PC?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
on mini mainframe ?	Y	N	Y	Y	N	N	N	N	N	N
type of calculations										
process design	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
mech. design	Y	Y	N	Y	Y	Y	N	N	N	N
rating	Y	Y	Y	Y	Y	Y	Y	Y	N	N
network design	Y	Y	Y		N	N	Y	N	Y	Y
network rating	Y	Y	Y		N	N	Y	N	N	N
other features										
physical	Y	Y			Y	Y	Y	N	N	N

property data										
cost estimating	Y	Y	N	Y	Y	N		N	Y	N
other heat-transfer eq.	Y	Y	N	Y	N	N	N	Y	N	N
integratable	Y	Y	N		N	N	N	N	N	N
with process simulation?										
pinch analysis	Y	Y							Y	

Distillation Columns. Distillation is by far the most common separation technique in the chemical process industries. Tray and packed columns are employed as strippers, absorbers, and their combinations in a wide range of diverse applications. Although the components to be separated and distillation equipment may be different, the mathematical model of the material and energy balances and of the vapor–liquid equilibria are similar and equally applicable to all distillation operations. Computation of multicomponent systems are extremely complex. Computers, right from their earliest availabilities, have been used for making plate-to-plate calculations.

A wide array of general-purpose distillation packages are available to the engineer. Some of the distillation software is stand-alone, whereas other packages are a part of a general-purpose flow sheet or process-simulation system. Because distillation is so universal, all process simulators have one or more distillation program modules for this unit operation. Often the nature of the distillation modules determines the suitability of or the preference for the use of a specific simulator for an application.

Distillation applications can be characterized by the type of materials separated, such as petroleum applications, gas separations, electrolyte separations, etc. These applications have specific characteristics in terms of the way or the correlations by which the physical properties are determined or estimated; the special configurations of the process equipment such as having side strippers, multiple product withdrawals, and internal pump arounds; the presence of reactions or two liquid phases; etc. Various distillation programs can model these special characteristics of the applications to varying degrees and with more or less accuracy and efficiency.

In petroleum refining, crude oil is separated into a number of fractions. The boiling range of components in crude oil is from 10° to 800°C, and for the purpose of distillation calculations, crude oil is represented as consisting of 50 to 100 narrow boiling fractions. The crude oil towers have many side streams, such as gasoline, kerosene, heating oil, etc, and there are associated steam strippers through which these side streams go. The computer programs that can handle the crude unit simulation must be able to deal with these special, complex characteristics.

In chemical and petrochemical applications, the chemical species are well defined for determination of their physical properties. For some chemicals, special methods of estimation may be needed, and certain mixtures may exhibit nonideal behavior; some even form azeotropes. The distillation program may not converge to a solution because of strong nonlinearities in the equations for the physical properties estimations arising from this nonideal behavior. Presence of electrolytes in the system also introduces peculiar nonlinearities and reactions, and some distillation programs may similarly not be able to handle them. At times, applications such as separation of isomers involve narrow boiling mixtures. In such cases, the separation requires many trays or stages, the calculations become lengthy, and some programs may not be able to handle them in a reasonable amount of time. At times such lengthy calculations can cause numerical problems and make convergence impossible.

Almost all distillation programs require the number of theoretical trays as input, and they calculate the separation that would emanate under various conditions of reflux and restriction. In this sense, they are all rating programs. The design determination of the number of theoretical stages is made by running the program as a case study, assuming various values of the number of stages, and then picking the one that meets all the requirements for the separation.

Because many variables are generally involved in a distillation calculation, it is necessary to see the profiles of flows, temperatures, pressures, and compositions inside the column before an engineer can be sure of the design of the column. It also is necessary to make parametric studies for some of the inputs such as heat duties, reflux ratios, feed locations, etc, to optimize the design. Most of the programs provide the profiles as output, and there is a trend toward displaying this information graphically. Similarly, the effect of the parametric variations also is shown graphically in some programs.

Once the process design of a distillation column is completed, the column diameter and column height can be determined using other programs or using special calculation options in some of the general distillation programs. Design of internal trays of a column is a more complicated procedure, and computer programs are available for that purpose. The mechanical design of the column requires other programs that consider the thermal and mechanical stresses of the vessel of the column and those on the internal trays.

Many industrial separations require a series of columns that are connected in specific ways. Some distillation programs can model such a system as a hypothetical single column with arbitrary cross-flows and connections and then carry out the distillation calculations for the modeled hypothetical column. Alternatively, such a system can be modeled as a process flow sheet using a process simulator.

Other Equipment. A large number of computer programs are available for design and rating of various types of process equipment, such as pumps, compressors, piping systems, valves, instruments, tanks, etc. Additional programs are being developed continually by manufacturers of processing equipment, manufacturers of chemicals, engineering contractors, and independent engineering software developers. Various trade associations and professional societies produce indexes and catalogs of such programs (eg, the AIChE software catalog). Many periodicals and newsletters carry regular columns announcing and describing such software (eg, *Chemical Engineering Progress* and *Chemical Engineering Magazine*).

Almost all vendors and manufacturers of process equipment use computer programs in designing and rating the equipment that they make. Such programs are made available to the purchasers of their equipment, often as a service or at a minimal cost. From an engineering point of view, some of these programs are biased in favor of the vendor's equipment.

Optimization

Finding the best solution when a large number of variables are involved is a fundamental engineering activity. The optimal solution is with respect to some critical resource, most often the cost (or profit) measured in dollars. For some problems, the optimum may be defined as, eg, minimum solvent recovery. The calculated variable that is maximized or minimized is called the objective or the objective function.

Exhaustive enumeration is the most straightforward method of finding an optimum when the number of possibilities is limited and manageable with the computer power available to an engineer. Many times, certain pieces of equipment such as heat exchangers, reactor vessels, etc, come in discrete sizes. In such cases, the objective function can be calculated for each combination of possible variables, and the optimum can be found. The computer makes it possible to consider many more combinations than otherwise possible. Many stand-alone design programs and most process simulators have a way of making and recording case studies, varying certain parameters so that the optimum can be found by the user conveniently.

In the design of a distillation column for a given separation, increasing the reflux ratio would reduce column height and increase its diameter and the operating cost for the column. An optimum reflux ratio can be found for the minimum total cost. In designing or developing a process with a reactor, separator, and recycle of unreacted materials, the engineer can optimize the design and size of the reactor and determine the size of the separator and the recycle constraints.

Many process simulators come with optimizers that vary any arbitrary set of stream variables and operating conditions and optimize an objective function. Such optimizers start with an initial set of values of those variables, carry out the simulation for the entire flow sheet, determine the steady-state values of all the other variables, compute the value of the objective function, and develop a new guess for the variables for the optimization so as to produce an improvement in the objective function.

There are several mathematical methods for producing new values of the variables in this iterative optimization process. The relation between a simulation and an optimization is depicted in Figure 6. Mathematical methods that provide continual improvement of the objective function in the iterative

process can be classified into the following categories: mountain-climbing procedures, iterative uses of linear programming, methods using penalty functions, and successive quadratic programming.

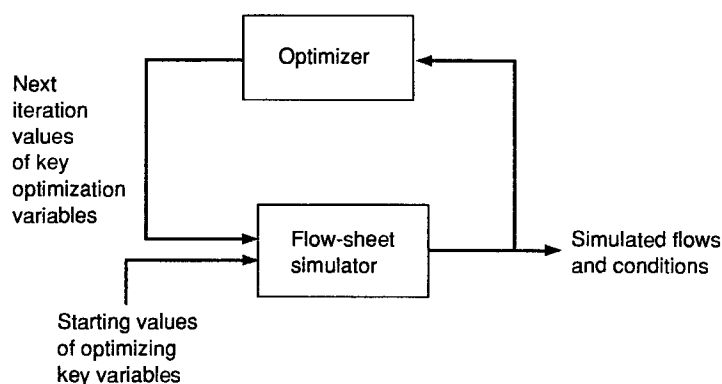


Fig. 6. The relation between a simulation and an optimization.

Mountain-Climbing Procedures. Mountain climbing is a method that is common, general, and intuitive (42). In this method, one solution for the entire flow sheet is obtained for a given set of the variables to be optimized and the value of the objective function is determined. The values of the variables are then perturbed or changed either by a small amount in their absolute value or by a small percentage of the values. The flow sheet is calculated for this new set of values and the objective function is recalculated. If the changes in the values and the objective function is an improvement in the objective function, then the new values are taken as a starting point, and perturbation in the same direction as those made in the previous step are made again; the process is continued as long as the objective function keeps improving. If the perturbation, on the other hand reduces the value of the objective function, then new perturbations are made in the direction opposite to the previous ones. If this perturbation produces an improvement in the objective function, then the process is continued.

When an optimum is reached, the perturbation in any direction would reduce the value of the objective function. Such an optimum, however, does not guarantee a global optimum, and the mountain-climbing process stops at a local optimum.

There are many methods of defining the set of new values of the variables that might be taken as each iteration in this mountain-climbing process. One of the earliest methods of determining the direction of the new values of the variables involves a search in random directions (43,44). Several vectors of random directions in the space of the variables of optimization are obtained, and the one showing the best improvement in the objective function is chosen for selecting the next point from which the process is repeated. At each stage of this process, the number of random directions to be evaluated can be varied, and the distance at which the objective function may be evaluated in any direction can be varied. The computation required for convergence to optimum depends on parameters such as these.

Once a direction is established for the next point in the space of the variables of optimization (whether by random search, by systematic evaluation of gradients, or by any other methods of making perturbations), it is possible to take a jump in the direction of the improvement much greater than the size of the perturbations. This could speed up the process of finding the optimum and reduce computer time. If such a leap is successful, the next iteration may take a bigger leap and so on, until the improvement stops. Then one could reverse the direction and decrease the size of the step until the optimum is found.

The perturbation in the values of the variables may be carried out one at a time, and thereby the partial derivatives of the objective function may be determined with respect to each variable. These in turn would enable one to find the gradient, which is the direction of the maximum rate of change of the objective function. The search for the optimum may then be continued along this direction. The gradient itself may be recomputed on the basis of one of many possible situations or rules. In the gradient-search method, the gradient is evaluated after each step (45).

Certain pathological objective functions create difficulties with various methods of mountain climbing. Special algorithms have been developed for such situations.

Iterative Use of Linear Programming. Linear programming is a powerful and computer-efficient method for finding an optimum for linear objective functions and linear constraints. When a situation can be modeled approximately as a linear model, then computer codes available for general linear programming can be used and the optimum can be found readily. For a special case of linear objective function and for the variables, all of which have linear equalities and inequalities (known as constraints), efficient methods of linear programming have been developed for finding an optimum.

For continuous variables without constraints, optimum values can be found mathematically by using advanced calculus. Numerically, such optimum values can be found by computer programs, a number of which are available with different degrees of sophistication (46,47).

In real-life problems in the process industry, nearly always there is a nonlinear objective function. The gradients determined at any particular point in the space of the variables to be optimized can be used to approximate the objective function at that point as a linear function; similar techniques can be used to represent nonlinear constraints as linear approximations. The linear programming code can then be used to find an optimum for the linearized problem. At this optimum point, the objective can be reevaluated, the gradients can be recomputed, and a new linearized problem can be generated. The new problem can be solved and the optimum found. If the new optimum is the same as the previous one then the computations are terminated. Otherwise, the iterations consisting of formulation of linear approximations and the solution of the linear programming problem are continually repeated (48).

This method of optimization is known as the generalized reduced-gradient (GRG) method. The objective function and the constraints are linearized in a piecewise fashion so that a series of straight-line segments are used to approximate them. Many computer codes are available for these methods. Two widely used ones are GRGA code (49) and GRG2 code (50).

Methods Using Penalty Functions. The essential element of using penalty methods is to convert a constrained optimization problem into an unconstrained optimization problem by modifying the objective function in such a way that the solution to this unconstrained problem will turn out to be the same as the solution to the original constrained problem. This is done by adding penalty terms to the objective function. As the value of the variable approaches the constraint, the penalty begins to grow. An example of a simple penalty function for an impenetrable constraint might be the reciprocal of the distance from the constraint, which would grow to large values as the variables approach the improbable constraint. Once the unconstrained optimization problem has been formulated, the optimization can be obtained by using any of the hill-climbing methods described above.

The generation of the penalty function requires selection from among many possible different functions as well as selection of many parameters of the penalty function. The contribution to the objective by the penalty function, relative to the contribution of the original objective function, changes the nature of the original objective function, which in turn influences the ease or difficulty of finding the optimum.

Successive Quadratic Programming. Successive quadratic programming recognizes the inherent nonlinearities of mathematical models of chemical processes through the use of quadratic expressions for local approximations of the objective function. The quadratic expression is used in contrast with linear expressions in the method of successive linear programming described above. The constraints continue to be approximated as linear constraints. Mathematical methods as computer codes have been developed as general-purpose programs for the problems thus formulated; these are known as quadratic programming.

This code is invoked for the process optimization problem once it is formulated as a quadratic problem locally. The solution from the code is used to arrive at the values of the optimization variables, at which the objective function is reevaluated and a new quadratic expression is generated for it. The

process is repeated just as in successive linear programming until a converged situation is found.

Successive quadratic programming is a refinement and enhancement of the quadratic programming, wherein the optimization problem is reformulated in terms of the direction of improvement from a starting set of values of the variables of optimization. There is a mathematical procedure for reformulating the constraints as linear approximated constraints in terms of the new variables, and the reformulated objective function becomes a quadratic function. Thus a quadratic programming problem is produced in the new variables, which define the direction from the starting point to the optimal point. The solution of the original optimization problem is now progressed along this direction as far as the objective function continues to improve.

At this point, the procedure of reformulating the gradient progressing problem in terms of the direction of improvement of the independent variables is repeated, and the operations described in the foregoing paragraph are carried out as the next iteration. The iterations are stopped when no further improvements can take place.

For optimization of process design and process control, the efficiency and effectiveness of the various methods depend on the process being modeled and the process modeling software that is used.

At times, it is possible to build an empirical mathematical model of a process in the form of equations involving all the key variables that enter into the optimization problem. Such an empirical model may be made from operating plant data or from the case study results of a simulator, in which case the resultant model would be a model of a model. Practically all of the optimization techniques described can then be applied to this empirical model.

Dynamic Models and Process Control

Mathematically speaking, a process simulation model consists of a set of variables (stream flows, stream conditions and compositions, conditions of process equipment, etc) that can be equalities and inequalities. Simulation of steady-state processes assume that the values of all the variables are independent of time; a mathematical model results in a set of algebraic equations. If, on the other hand, many of the variables were to be time dependent (in the case of simulation of batch processes, shutdowns and startups of plants, dynamic response to disturbances in a plant, etc), then the mathematical model would consist of a set of differential equations or a mixed set of differential and algebraic equations.

There are special numerical analysis techniques for solving such differential equations. New issues related to the stability and convergence of a set of differential equations must be addressed. The differential equation models of unsteady-state process dynamics and a number of computer programs model such unsteady-state operations. They are of paramount importance in the design and analysis of process control systems (see PROCESS CONTROL).

Data Reconciliation

The purpose of data reconciliation is to produce one set of consistent data that can be used for materials accounting, for monitoring the operations of a plant, and for determining the parameters of a model of a process unit. Process-plant information systems, the use of which is becoming widespread, produce a vast amount of process data with the possibility of a great deal of redundancy and replication. Because the data come from measuring instruments, there could be errors of measurements in the data. The data could be incomplete in the sense that some measurements may be missing, thus it is impossible to make a material or energy balance around a process unit or a plant or to determine a calculated parameter such as plate efficiency of a distillation column or fouling factor of a heat exchanger. On the other hand, data may be redundant in the sense that different selection of subsets of the data could give different material balances or produce different values of parameters or coefficients in the models of the process equipment.

Reconciled data must satisfy all material and energy balance constraints and those of rate equations for which coefficients and parameters are fully known. The mathematical process of reconciliation can be described schematically as in Figure 7.

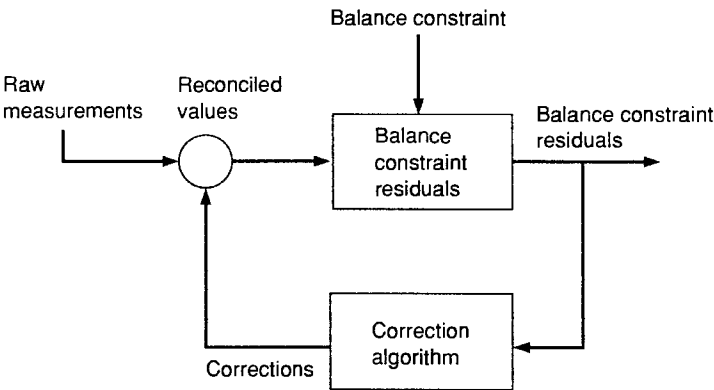


Fig. 7. Data reconciliation.

Raw data are repeatedly corrected by an amount determined by a correcting algorithm and checked against the constraints they must satisfy. The residuals of the constraints, which is a measure of the degree to which the constraints are not redefined, are calculated and the algorithm attempts to minimize these residuals. The procedure is continued until the residuals can no longer be reduced.

Process Synthesis

Process synthesis is the step in design when the chemical engineer selects component parts and the interconnection between them to create the flow sheet. This formal approach to design includes developing a representation of the synthesis problem, using a means to evaluate alternatives, and following a strategy to search the almost infinitely large space of possible alternatives. Effective solutions depend heavily on the nature of the synthesis problem being addressed (51,52).

Whereas process simulation includes quantitative analysis of a design given the structure of the design, process synthesis involves determining the structure that will meet the requirements of the design as well as finding the best structure for the requirements. For example, if components A, B, C, and D whose relative volatilities were in the order D, C, B, and A were to be separated by distillation for which each column produced a top and a bottom fraction, five schemes of three columns arise as possible structures (53) (Fig. 8).

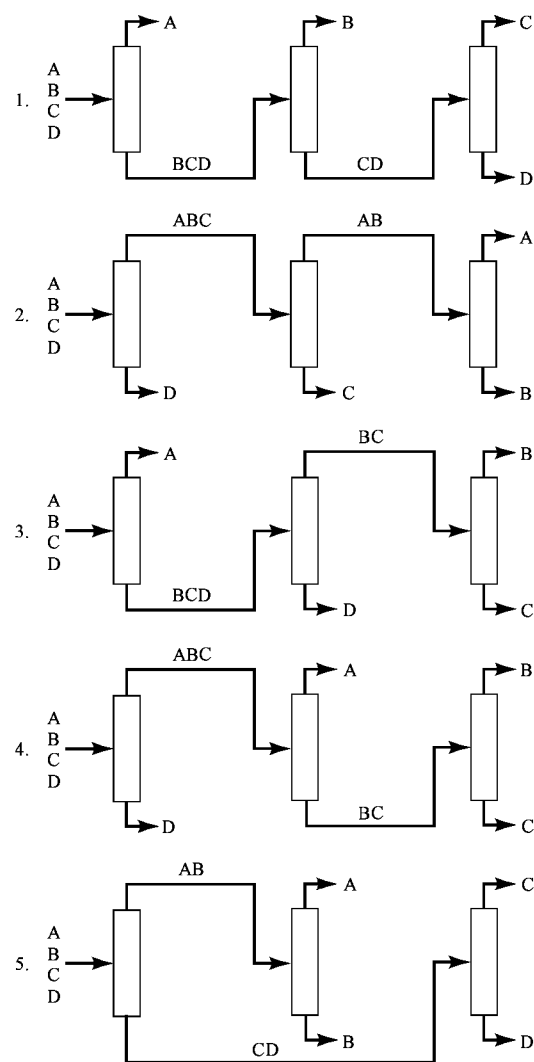


Fig. 8. Five schemes for a design structure for separating components A, B, C, D with volatilities in the order D, C, B, A.

The task of process synthesis is to evaluate these schemes and select the best. The creative possibilities of design do not stop here. For example, these separations can be carried out by heat integration, by multiple-draw columns, or by use of processes other than distillation; many other variations are possible.

Any formal approach to process synthesis of an industrial problem soon grows into a computational problem of developing an astronomical number of alternatives and then finding a way of selecting among them. The two main approaches to these problems are heuristic and mixed-integer programming. The heuristic approach involves making up different case studies and evaluating them. Mixed-integer programming is a mathematical tool wherein all the defined alternate structures are modeled as a mathematical programming problem and special computer codes produce an optimal solution with the structure and the values of all the variables that produce the best value of the objective function.

Mixed-integer programming contains integer variables with the values of either 0 or 1. These variables represent a structure or substructure. A special constraint about the structures states that of a set of (structure) integer variables only one of them can have a value of 1; expressed in a statement: the sum of the values of (alternate) variables is equal to 1. In this manner, the arbitrary relations between structures can be expressed mathematically and then the optimal solution is found with the help of a computer program. (52).

Typically, process synthesis involves the synthesis of the following: heat-exchanger networks, separation systems, chemical reaction paths, complete flow sheets, and control systems. Significant progress has been made in developing insights to aid in the design of heat-exchanger networks, and this area has been reviewed (54). Although synthesis of heat-exchanger networks can be used for grass-roots design of new chemical plants, it also provides a systematic technique for optimal retrofit of existing heat-exchanger networks. The tasks involved are

- (1) Selecting the process stream matches that will participate in the retrofitted network and their heat loads.
- (2) Determining the retrofitted network configuration.
- (3) Calculating the heat-transfer area of any new exchangers and for retrofitting.
- (4) Deciding whether a match will be housed in its original exchanger, a different exchanger, or in a new exchanger.
- (5) Obtaining the amount of area that must be added to an existing exchanger.

These tasks are interrelated. One cannot derive a network before selecting process stream matches or determine the best-match exchanger assignments without knowing the heat-transfer area requirements of each match. There are many solutions to tasks 1 and 2. For any given retrofit problem, there are several potential combinations of process stream matches. Each of these combinations has several potential network configurations, and for each network configuration, there are several ways to assign the existing exchangers. The modification cost of the retrofitted network is determined by the streams that must be repiped, the number and size of any new exchangers, the amount of additional area that must be added to existing exchangers, and the costs associated with moving any existing exchangers. Thus the modification cost is determined by the results of tasks (42,44). Task 1 contributes indirectly to the cost of the retrofitted network, as the network configuration and heat-transfer areas will vary with the selected combination of matches and their heat loads.

The best way to approach the retrofit synthesis of the heat-exchanger network is to model all five tasks simultaneously. A mixed-integer nonlinear programming model is usually formulated to accomplish this goal.

A useful concept in heat-exchanger network design has been the identification of the pinch point of heat integration. A key assumption of many pinch-based design techniques has been that no heat can flow across the pinch, which results in a partitioning of the heat-exchanger network synthesis into subnetwork problems that can be solved independently. The pinch-point concept has been used as an alternate method to retrofit heat-exchanger networks (55). This procedure consists of (1) the identification of cross-pinch exchangers; (2) the elimination of the cross-pinch exchangers from the existing network; (3) the positioning of new exchangers and, where possible, the reuse of exchangers removed in step 2; and (4) the evolution of improvements by considering heat-load loops between streams and process and utility exchangers. However, such an approach can provide only bounds

within which it is expected to determine a good retrofit, and the best retrofit design is difficult due to the complication of process constraints and plant layout.

Synthesis of separation systems has mainly been confined to sequencing distillation columns by taking advantage of heat-integration options. For example, a simple synthesis method has been developed that is based on utility bounding, assuming that $Q \Delta T$, the product of the condenser or reboiler heat duty and the temperature difference between the reboiler and condenser is constant for each distillation column (56). Generally, the synthesis of heat-integrated distillation sequences is formulated as a mixed-integer linear or nonlinear programming problem.

Some progress has been made in the development of computer aids by chemists for reaction path synthesis, leading to desired complex organic molecules. Synthesis in areas such as complete flow sheeting and control systems is not industrially significant as yet.

Computer-aided process synthesis systems do not mean completely automated design systems (57). Process synthesis should be carried out by interactive systems, in which the engineer's role is to carry out synthesis and the machine's role is to analyze the performance of synthesized systems. Computer applications in the future will probably deal with the knowledge-based system in applied artificial intelligence. Consequently, research on computer-aided process synthesis should be directed toward the realization of such systems with the collaboration of experienced process engineers.

One approach relies heavily on heuristics but allows the engineer to interact during the synthesis procedure through a framework of hierarchical decision levels (58,59).

- Level 1: Batch versus continuous
- Level 2: Input-output structure of the flow sheet
- Level 3: Recycle structure of the flow sheet
- Level 4: Separation system specification
 - Level 4a: Vapor recovery system
 - Level 4b: Liquid recovery system
- Level 5: Heat exchanger network

This approach attempts to permit the designer to evaluate and optimize alternatives quickly, using shortcut hand calculations. However, if it is to be adopted by practicing design engineers it must be complemented by software for at least the analysis, evaluation, and optimization stages.

A hierarchical design procedure for process synthesis can be used in conjunction with a flow-sheeting program to analyze, evaluate, and optimize the options (60). The emphasis is on starting with the simplest possible models that will give answers to a particular question quickly so that the questions to be asked at the next decision level can be formulated. At each stage, it is necessary to ensure that the level of detail in the model is sufficient to give reliable information.

Artificial Intelligence and Expert Systems

Computers that started out as machines to carry out programmed calculations, were soon used as information-processing machines. The next progression was to use them for processing nonnumerical, nonquantitative information and program them so that they would simulate the way a human mind processes such nonquantitative information. Figure 9 illustrates this progression graphically.

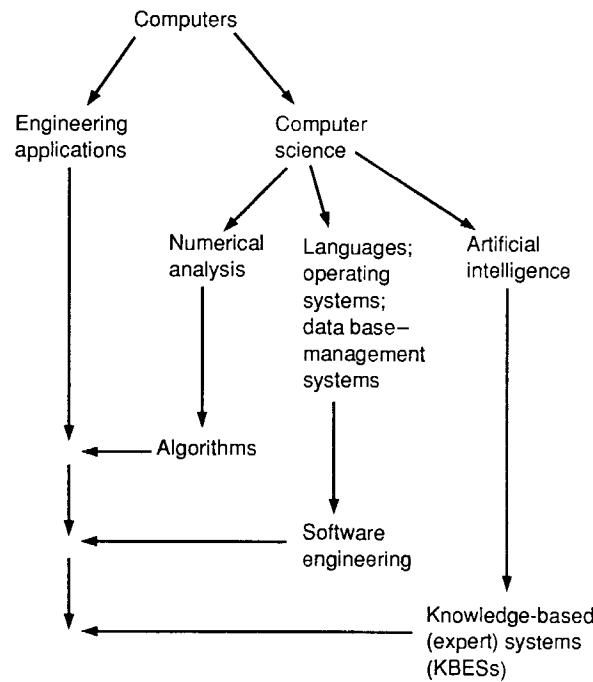


Fig. 9. Processing of nonquantitative information.

One of the early motivations for artificial intelligence came in the mid-1960s from the perceived potential for using computers to aid unstaffed space vehicles and crewless vehicles for exploring the surface of the moon. Special branches of studies of artificial intelligence such as artificial vision, pattern recognition, and robotics were identified and developed (61). An approach to working with nonquantitative symbolic information in the form of logical and linguistic statements using a computer developed into a software discipline known as expert systems.

The earliest practical use of an expert system was made in the software named MYCIN for diagnosing a toxic poison from the symptoms of a patient and recommending the antidote (62). This type of activity is generally carried out by a human expert who processes information about a situation (in this case, symptoms of a patient), refers to the expert's experience and expert knowledge, and then recommends action (in this case, the antidote).

Such nonquantitative application of computers is being broadened to include more creative activities such as design and planning, using some of the methodology and the software of expert systems. At times, such enhanced systems are given the generic name of knowledge-based systems. Figure 10 shows the elements of such a knowledge-based (expert) system.

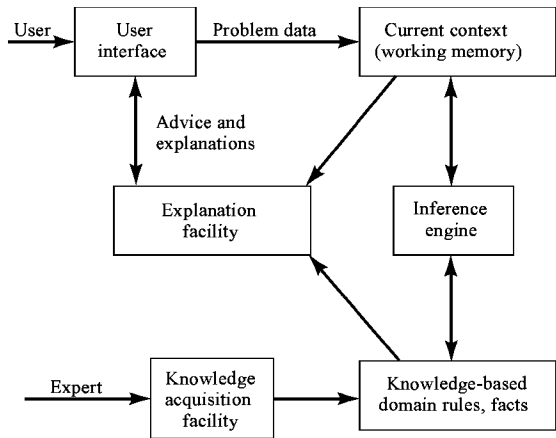


Fig. 10. Elements of an expert system.

Expert systems (qv) require a depository of relevant facts (known by experts) for the field or discipline the system is set up for. Most of the facts are in the form of IF ... THEN conditional statements, known as rules. For example: if the poison is *A* and elapsed time since taking poison is *Y* hours, then the symptom will be *B*. Software known as an inference engine is capable of using the library of facts and rules, of carrying out logical inferences forward and backward, of relating facts and rules relevant to a user's input inquiries, and of reporting deductions. The software deploys and stays within the confines of all rules of logic (formally known as propositional and predicate calculus). The system must deal with vast volumes of information and complex linguistic representations of such information and inquiries for practically useful applications.

The expert's knowledge must be incorporated in the expert system in a formal way such that it is unambiguous to the system and the system can make forward and backward inferences. This requires special training and skill, and the person possessing them is known as a knowledge engineer.

The earliest and most frequent use of expert systems so far has involved searching for information and searching for causes of situations, ie, applications involving diagnosis, interpretation, and monitoring. The first engineering application was the selection of the software and method for use in finite element analysis of stresses in a mechanical structure (63). An expert system called *ACONPHYDE* was developed for determining methods of estimating physical properties that can be used in specific situations, depending on the type of information available about compounds or mixtures (64). A system called *PROSPECTOR* helps in interpreting geological surface data to evaluate potential mineral deposits (65). In the process industry, applications have developed in diagnosing faults in operating processes, from the analysis of the processes, from examination of the operating procedures, and from monitoring of the operating data.

The early expert systems such as MYCIN typically involved 5 to 10 researchers for 2 to 5 years and cost several million dollars (66). Improved software tools that include expert diagnostic shells, interactive graphics, and object-oriented programming have reduced this effort by a factor of 10 to 100, and many software shells and other programs are now available on microprocessors and desktop computers (67). For many classes of engineering applications involving diagnosis, evaluation, monitoring, or configuration, there now exist knowledge-based expert system shells and tools whose representation and reasoning paradigms match the requirements of applications (66). In any application, a considerable amount of knowledge engineering work still needs to be carried out.

In addition to the knowledge based on the experience of human experts, the expert system might contain information about fundamental laws of physics, chemistry, or other sciences. This could be in the form of a mathematical model of a process. The system could then be expected to make inferences based on this fundamental knowledge. The foregoing two forms of knowledge are known as shallow knowledge and deep knowledge, respectively. The inference engine of the expert system produces additional rules based on inferences made from the deep knowledge, which are then called compiled or derived knowledge. To deal with deep knowledge, large computer memory and much computer time are required. Currently, there are no mature or industrially significant systems in existence that operate at that level.

The process of design, including the design of chemical processes, is a complex nonquantitative process requiring a vast amount of information, deep knowledge of the field, and expert experience. Artificial intelligence will most likely be applied in developing process designs in the future. This is a promise and a possibility that is a target of current research (68).

Fault Tree Analysis

Fault tree analysis (FTA) is a widely used computer-aided tool for plant and process safety analysis (69). One of the primary strengths of the method is the systematic, logical development of the many contributing factors that might result in an accident. This type of analysis requires that the analyst have a complete understanding of the system and plant operations and the various equipment failure modes.

FTA breaks down an accident into its contributing equipment failures and human errors (70). The method therefore is a reverse-thinking technique, ie, the analyst begins with an accident or undesirable event that is to be avoided and identifies the immediate cause of that event. Each of the immediate causes is examined in turn until the analyst has identified the basic causes of each event. The fault tree is a diagram that displays the logical interrelationships between these basic causes and the accident.

The result of the FTA is a list of combinations of equipment and human failures that are sufficient to result in the accident (71). These combinations of failures are known as minimal cut sets. Each minimal cut set is the smallest set of equipment and human failures that are sufficient to cause the accident if all the failures in that minimal set exist simultaneously. Thus a minimal cut set is logically equivalent to the undesired accident stated in terms of equipment failures and human errors.

The following symbols are used in fault tree construction to display the interrelationships between equipment failures and a specific accident:

- The *TOP event* is the specific accident or undesired event that is the subject of the FTA.
- The *OR gate* indicates that the output event occurs if any of the inputs occur.
- The *AND gate* indicates that the output event occurs only when all the input events occur.
- The *INHIBIT gate* indicates that the output event occurs when the input event occurs and the inhibit condition is satisfied.
- The *DELAY gate* indicates that the output event occurs when the input event has occurred and the specified delay time has expired.
- The *BASIC event* represents a basic equipment fault or failure that requires no further development into more basic faults or failures.
- The *INTERMEDIATE event* represents a fault event that results from the interactions of other fault events that are developed through logic gates such as those defined above.
- The *UNDEVELOPED event* represents a fault event that is not examined further because information is unavailable or because its consequence is insignificant.
- The *EXTERNAL or HOUSE event* represents a condition or an event that is assumed to exist as a boundary condition for the fault tree.
- The *TRANSFER IN symbol* indicates that the fault tree is developed further at the occurrence of the corresponding *TRANSFER OUT symbol*.

Fault Tree Construction. Fault tree construction begins at the TOP event and proceeds, level by level, until all fault events have been developed to their basic contributing causes (BASIC events). The analyst begins with the TOP event and, for the next level, determines the immediate,

necessary, and sufficient causes that result in the TOP event. Normally, these are not basic causes but are intermediate faults that require additional development. If the analyst can immediately determine the basic causes of the TOP event, the problem is not suitable (too simple) for fault tree analysis and can be evaluated by other methods such as failure modes, effects, and criticality analysis (FMECA). Figure 11 shows a representative fault tree structure.

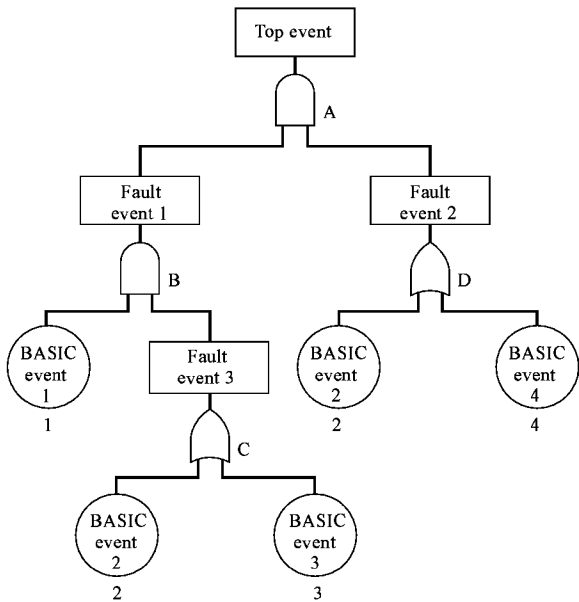


Fig. 11. A fault tree structure.

The immediate causes of the TOP event are shown in the fault tree with their relationship to the TOP event. If any one of the immediate causes results directly in the TOP event, the causes are connected to the TOP event with an OR logic gate. If all the immediate causes are required for TOP event occurrence, then the causes are connected to the TOP event with an AND logic gate. Each of the immediate causes is then treated in the same manner as the TOP event, and its immediate, necessary, and sufficient causes are determined and shown on the fault tree with the appropriate logic gate. This development continues until all intermediate fault events have been developed into their basic causes.

Fault Tree Solution. Solving the fault tree means obtaining the minimal cut sets. The minimal cut sets are all the combinations of equipment failures that can result in the fault tree TOP event. Computer programs are required to determine the minimal cut sets for large fault trees (72). The solution method has four steps:

- (1) Uniquely identify all gates and BASIC events.
- (2) Resolve all gates into BASIC events.
- (3) Remove all duplicate events within sets.
- (4) Delete all supersets (sets that contain another set as a subset).

Minimal cut sets are then ranked. Two factors are considered in the ranking procedure. The first factor considers structure, ie, a one-event minimal cut set is more important than a two-event minimal cut set. The implication is that one event is more likely to occur than two events, two events are more likely than three events, and so on. The second factor considers ranking within equal-size minimal cut sets. The general ranking rules consider the probability of human error, active equipment failure, and passive equipment failure (73).

This ranking implies that human errors are more likely to occur than active equipment failures (functioning equipment, such as a running pump) and that active equipment failures are more likely to occur than passive equipment failures (static, nonfunctioning equipment, such as a storage tank).

Spreadsheet Programs

Spreadsheet programs are used by engineers to perform a variety of computer design and analysis functions. The applications that are implemented on them include process design and optimization, equipment design, batch distillation, process control, process economics, and many other functional areas of interest. They have the advantage of making it easy to mount an application, as opposed to implementing it as a special-purpose program written in some higher level language such as FORTRAN, C, Pascal, etc. Naturally, not every application is suited to a spreadsheet; for example, they are difficult to apply to calculations involving nonlinear physical properties and highly nonideal vapor–liquid equilibria. But they come into their own in the ease with which a complex design problem can be investigated, professional-looking graphic output can be obtained, parametric analyses can be done, etc.

Spreadsheet programs were originally developed for bookkeeping and business-planning applications; over time, many additional features were added. *Visicalc*, which appeared in the early 1980s, was the first spreadsheet program; it was developed for Apple computers.

A spreadsheet program is intuitive in its operation and can immediately show the effect of any one change throughout the whole spreadsheet. It is thus a subtle form of data validation; an error may be spotted immediately. Furthermore, when a material balance is set up, the total effect caused by a change in one variable may be seen at once.

Basically, a spreadsheet program is one that operates on many rows and columns (8196 rows and 256 columns is currently the minimum for any respectable program). A caption or a value may be placed in any row–column location (which is usually referred to as a cell); alternatively, a cell may contain a formula for how its value is to be computed as a function of values stored in other cells. Rows or columns (or portions thereof) may be summed or be involved in other mathematical calculations. Logic tests for conditional results can be put into a cell, regression analyses can be performed on a series of cells, averages can be computed, and many other mathematical functions (rounding, exponentiation, calculation of net present value, etc) are available. A series of cells may also be plotted graphically in many ways (line, bar, pie charts, etc) so that presentation of results is simplified for the user while giving the output a professional appearance. Among the available spreadsheet programs, there is *Lotus 1-2-3* (Lotus Development Corp.), *Symphony* (Lotus Development Corp.), *Full Impact* (Ashton Tate), *Superak* (Computer Associates), *Excel* (Microsoft Corp.), and *Quattro Pro* (Borland International, Inc.).

Another implicit advantage to use of spreadsheets, besides their ubiquity and ease of use, is the fact that so many engineers and scientists already are familiar with them. Their use then obviates the need to learn another programming language.

Spreadsheet Applications. The types of applications handled with spreadsheets are a microcosm of the types of problems and situations handled with full-blown application programs that are run on microcomputers, minis, and mainframes and include engineering computations, process simulation, equipment design and rating, process optimization, reactor kinetics–design, cost estimation, feedback control, data analysis, and unsteady-state simulation (eg, batch distillation optimization).

In using a spreadsheet for process modeling, the engineer usually finds it preferable to use constant physical properties, to express reactor performance as a constant "conversion per pass," and to use constant relative volatilities for distillation calculations; such simplifications do not affect observed trends in parametric studies and permit the user quickly to obtain useful insights into the process being modeled (74,75).

When an engineer sets up the model of a process with a recycle stream, at some point in the problem formulation, he or she will come across a cell (ie, variable) whose value depends on itself. That is perfectly normal. Starting with an initial guess for the variable (which may be zero), a value is calculated for that variable, and this calculated value is subsequently used as the initial guess for the next round of calculations. This method of direct substitution (also known as the Gauss-Seidel method) occurs every time the spreadsheet is recalculated, and recalculations are under the direct control of the user in all spreadsheet programs via the use of a specially designated key. After a number of these recalculations (iterations), the difference between the initial guess and the calculated value approaches zero (ie, the values become identical), and convergence is reached. An added benefit is the ability to observe the sensitivity of all the other parameters in the spreadsheet as a function of the convergence process (74).

Addition of unit costs (or relative cost factors) permits either cost estimation or process optimization to be carried out via the spreadsheet program; the original use of spreadsheets was for business and financial analyses. The engineer can decide what it is to be optimized (ie, maximized or minimized), and if this is the net present value or internal rate of return of a series of periodically evaluated costs, these functions are already available for use within the program. Alternatively, if the engineer wishes to maximize a profit that is the result of a trade-off between equipment amortization costs and operating costs, the program can be directed to build a table of values from which the optimum can easily be observed (74).

A useful feature of spreadsheet programs is the ability to copy and replicate portions of the spreadsheet that have already been entered. This replication includes not only values but also processing rules that have been stored in each cell; during copying, the rules are adjusted to be relative to the cell being copied, if so specified. This makes the modeling of reaction kinetics a relatively simple matter; the reactor may be divided into *n* differential slices, and the differential equations governing the kinetics are entered into the spreadsheet only for the first slice. Once that slice has been set up and debugged, the user simply copies the relevant portion of the spreadsheet *n* times to obtain a spreadsheet model of the complete reactor (74).

Another important feature of spreadsheet programs is the ability to generate graphical output. Although the utility of such a feature is immediately obvious for the presentation of cost and economic information, cost comparisons, etc, its usefulness in a purely engineering context cannot be overlooked. For instance, problems relating to fluid-flow analysis, pump curves, head curves, pressure drops, etc, may be plotted as a function of flow rate, varying pipe size, etc, and the solution can be obtained by examining the point at which the two curves appropriately intersect. Visually examining the curves and noting their sensitivity to the extensive parameter can also prove useful (76,77).

Design and analysis programs involving the spreadsheet approach are available for a wide array of engineering problems. These programs are called templates, and they load the prepackaged spreadsheet into the spreadsheet program. The template contains the column and row headings, built-in constants, and built-in processing rules for all the relevant cells. The user may use the template "as is" or edit it to adapt it to the special situation at hand. If the template exactly fits the problem the user wishes to solve, only those cells that represent input variables are entered into the spreadsheet (74).

Current spreadsheet programs contain many advanced capabilities and enhancements that an engineer might want to use; they also offer improved performance, interfaces to data base and word-processing programs and have a host of other features. A good overview that describes the features and presents what is available in each of the latest spreadsheet packages on the market is available (78).

BIBLIOGRAPHY

1. C. G. Kirkbride, *Chemical Engineering Fundamentals*, McGraw-Hill Book Co., New York, 1947.

2. R. Annino and R. Driver, *Scientific and Engineering Applications with Personal Computers*, John Wiley & Sons, Inc., New York, 1986.

3. W. L. Luyben, *Process Modelling, Simulation and Control for Chemical Engineers*, McGraw-Hill Book Co., New York, 1990.

4. R. S. H. Mah, *Chemical Process Structures and Information Flows*, Butterworth & Co. (Publishers) Ltd., London, 1990.

5. L. T. Biegler, *Chem. Eng. Prog.* **85**(10), 50–61.

6. G. Adkins and U. W. Pooch, *Computers* **10**, 12 (1978).

7. D. A. Spitz, *Instrum. Technol.* **26**, 81 (1979).

8. J. J. Davis and R. I. Kermode, *Ind. Eng. Chem. Process Des. Dev.* **14**, 459 (1975).

9. B. W. Overturf, G. V. Reclaitis, and J. M. Words, *Ind. Eng. Chem. Process Des. Dev.* **17**, 161 (1978).

10. U. M. Diwekar and E. S. Rubin, *Comput. Chem. Eng.* **15**, 105–114 (1991).

11. S. S. Grover, A. M. Peiser, M. K. Sood, M. T. Tayyabkhan, and G. M. Weber, *Chem. Eng. Prog.*, **79**(6), 54–61 (1983).

12. A. E. Ravitz and R. L. Norman, *Chem. Eng. Prog.* **60**, 71 (1964).

13. M. T. Tayyabkhan, *paper presented at the AIChE National Meeting*, Cleveland, Ohio, May 1961.

14. O. Redlich and J. N. S. Kwong, *Chem. Rev.* **44**, 233 (1949).

15. G. Soave, *Chem. Eng. Sci.* **27**, 1197 (1972).

16. D. Y. Peng and D. B. Robinson, *Ind. Eng. Chem. Fundam.* **15**, 59 (1976).

17. B. I. Lee and M. G. Kesler, *AIChE J.* **21**, 510 (1975).

18. J. G. Hayden and J. P. O'Connell, *Ind. Eng. Chem. Fundam.* **14**, 209 (1975).

19. R. W. Hankinson and G. H. Romson, *AIChE J.* **25**, 653 (1979).

20. H. G. Rackett, *J. Chem. Eng. Data* **15**, 514 (1970).

21. G. Scatchard, S. E. Wood, and J. M. Mochel, *J. Am. Chem. Soc.* **68**, (1957, 1946).

22. A. Fredenslund, R. L. Jones, and J. M. Prausnitz, *AIChE J.* **21**, 1086 (1975).

23. D. S. Abrams and J. M. Prausnitz, *AIChE J.* **21**, 116 (1975).

24. G. M. Wilson and C. H. Deal, *Ind. Eng. Chem. Fundam.* **1**, 20 (1962).

25. G. M. Wilson, *J. Am. Chem. Soc.* **86**, 127 (1964).

26. H. Renon and J. M. Prausnitz, *Ind. Eng. Chem. Process Des. Dev.* **8**, 413 (1969).

27. K. Lucas, *Chem. Eng. Technol.* **46**, 157 (1979).

28. T. H. Chung, L. L. Lee, and K. E. Starling, *Ind. Eng. Chem. Fundam.* **23**, 8 (1984).

29. D. Reichenberg, Report 38, Division of Chemical Standards, National Physics Laboratory, UK, Aug. 1975.

30. J. F. Ely and H. J. M. Hanley, *Ind. Eng. Chem. Fundam.* **20**, 323 (1981).

31. D. Van Velzen, R. L. Cardozo, and H. Lagenkamp, *Ind. Eng. Chem. Fundam.* **11**, 20 (1972).

32. J. W. Prevezdziecki and R. Sridhar, *AIChE J.* **31**, 333 (1985).

33. L. Grunberg and A. H. Nissan, *Ind. Eng. Chem.* **42**, 885 (1950).

34. A. S. Teja and P. Rice, *Ind. Eng. Chem. Fundam.* **20**, 77 (1981).

35. D. Roy and G. Thodos, *Ind. Eng. Chem. Fundam.* **9**, 71 (1970).

36. G. Latini, *Chem. Eng. World* **17**, 49 (1982).

37. R. L. Rowley, *Chem. Eng. Sci.* **37**, 897 (1982).

38. P. D. Neufeld, A. R. Janzen, and R. A. Aziz, *J. Chem. Phys.* **57**, 1100 (1972).

39. C. R. Wilke and C. Y. Lee, *Ind. Eng. Chem.* **47**, 1253 (1955).

40. C. R. Wilke and P. Chang, *AIChE J.* **1**, 264 (1955).

41. M. T. Tyn and W. F. Calus, *J. Chem. Eng. Data* **20**, 106 (1975).

42. V. E. Jenson and G. V. Jeffreys, *Mathematical Methods in Chemical Engineering*, Academic Press, Inc., Orlando, Fla., 1990.

43. R. J. Kelly and R. F. Wheeling, *A Digital Computer Program for Optimizing Non-linear Functions*, Mobil Oil Corp., Research Dept., Central Research Division, Princeton, N.J., July 1962.

44. R. Mihail and G. Maria, *Comp. Chem. Eng.* **10**, 539 (1986).

45. H. E. Zellnick, N. E. Sondak, and R. S. Davis, *Chem. Eng. Prog.* **58**(8), 35 (1962).

46. T. F. Edgar and D. M. Hinnelblau, *Optimization of Chemical Processes*, McGraw-Hill Book Co., Inc., New York, 1988.

47. G. V. Reklaitis, A. Ravindran, and K. M. Ragsdell, *Engineering Optimization—Methods and Applications*, John Wiley & Sons, Inc., New York, 1983.

48. F. Palacios-Gomez, L. Lasdan, and M. Engquist, *Mgt. Sci.* **28**, 1106 (1982).

49. J. Abadie and J. Guigou, *Gradient réduit généralisé, Electricité de France*, Note HI **069/02** Apr. 15, 1969.

50. L. S. Lasdon, N. Kim, and J. Zhang, *An Improved Successive Linear Programming Algorithm*, University of Texas, Austin, 1983.

51. N. Nishida, G. Stephanopoulos, and A. W. Westerberg, *AIChE J.* **27**, 321 (1981).

52. D. F. Rudd, G. F. Powers, and J. J. Sürola, *Process Synthesis*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1973.

53. F. Kaisel in *I Chem Energy Symposium Series* **109**, 1989, pp. 43–59.

54. T. Gundersen and L. Naess in *Proceedings of Chemical Engineering Fundamentals, CEF '87: The Use of Computers in Chemical Engineering*, 1987, pp. 675–704.

55. T. N. Tjoe and B. Linnhoff, *Chem Eng.* **93**, 47–60 (1986).

56. M. J. Andreconth and A. W. Westerberg, *AIChE J.* **31**, 363 (1985).

57. T. Uneda, *Comput. Chem. Eng.* **7**, 279 (1983).

58. J. M. Douglas, *AIChE J.* **3**, 353 (1985).

59. B. Carnahan, M. A. Luthur, and J. O. Wilkes, *Applied Numerical Methods*, John Wiley & Sons, Inc., New York, 1969.

60. D. H. Lott in *I Chem Energy Symposium Series* **109**, 1989, pp. 1–59.

61. *Artificial Intelligence*, Time-Life Books, Inc., Chicago, Ill., 1986.

62. B. G. Buchanan and E. H. Shortliffe, *Rule-Based Expert Systems*, Addison-Wesley Publishing Co., Inc., Reading, Mass., 1984.

63. J. S. Bennett and R. S. Englemore, in *Proceedings of AAAI—84*, 1984.

64. R. Banares-Alcantara, A. W. Westerberg, and M. D. Rychener, *Development of an Expert System for Physical Property Predictions, Technical Report*, Carnegie Mellon University, Design Research Center and Robotics Institute, Pittsburgh, Pa., 1983.

65. A. N. Campbell, V. Hollister, R. O. Duda, and P. E. Hart, *Science* **97**, 2173 (1982).

66. C. L. Dym and R. E. Levitt, *Knowledge-Based Systems in Engineering*, McGraw-Hill Book Co., Inc., New York, 1991.

67. L. Egol, *Chem. Eng.* **97**, 157 (Mar. 1990).

68. G. Stephanopoulos, *Comput. Chem. Eng.* **14**(11), 1259–1270 (1990).

69. American Institute of Chemical Engineers, *Guidelines for Hazard Evaluation Procedures*, American Institute of Chemical Engineers, Center for Chemical Plant Safety, 1985.

70. J. B. Fussell in *Generic Techniques in System Reliability Assessment*, Office of Nuclear Regulatory Research, Nuclear Regulatory Commission, Washington, D.C., 1976.

71. W. E. Vesely, *Fault Tree Handbook*, NUREG-0492, 1981.

72. J. B. Fussell, E. B. Henry, and H. N. Marshall, *MOCUS: A Computer Program to Obtain Minimal Cut Sets from Fault Trees*, Report 1156, Aerojet Nuclear Co., Idaho Falls, Idaho, 1974.

73. N. J. McCormick, *Reliability and Risk Analysis*, Academic Press, Inc., New York, 1981.

74. F. M. Julian, *Chem. Eng. Prog.* **85**(10), 33 (Oct. 1989).

75. E. M. Rosen and R. N. Adams, *Comp. Chem. Eng.* **11**, 723.

76. J. D. Kumana, *Chem. Eng. Prog.* **86**(12), 53 (Dec. 1990).

77. J. W. Holmes, *Chem. Eng.* **95**, 166 (Dec. 19, 1988).

78. *Chem. Eng. Prog.* (1990).

M. T. Tayyabkhan
Tayyabkhan Consultants, Inc.

H. Britt
Aspen Technology, Inc.

COMPUTER-AIDED PROCESS DESIGN.

See SIMULATION AND PROCESS DESIGN.

COMPUTER INTEGRATED PROCESSING.

See PROCESS CONTROL.

COMPUTER NETWORKS.

See COMPUTER TECHNOLOGY.

COMPUTER TECHNOLOGY

In the last 10 years the computer industry has experienced dramatic and fundamental changes. Computer users have witnessed both vastly increased capabilities and dramatically reduced costs. Chemists have often been at the forefront of taking advantage of the new capabilities. It seems likely that the next few years hold the promise of even greater rates of change and ever-expanding capabilities. The challenge of staying current with the technology will be ever more difficult.

The years since publication of the third edition of the *Encyclopedia* (1978–1984) have brought the rise and fall of the minicomputer, the worldwide ascendancy of microprocessor-based personal computers, the emergence of powerful scientific work stations, the acceptance of scientific visualization, further advances with supercomputers, the rise and fall of the minisupercomputer, and the realization that the future lies in parallel computing.

Each of these and other phenomena could, by themselves, benefit from in-depth examination. This article focuses primarily on those computing technologies that find application in computational domains, especially within computational chemistry.

Personal Computers

This sentence from *ECT* 3rd ed has proven prophetic:

An important recent development has been introduction of the microprocessor, or computer on a chip.

Microprocessor-based personal computers have since become ubiquitous in modern business practice, forever setting aside many previously accepted notions of the office. Not surprisingly, the principal gains went first to general business users, with spreadsheets and word processors becoming standard equipment. Over time, significantly faster computers have become available at ever lower prices. Thousands of software packages are available. Today there are more than 50 million personal computers in the United States alone; a \$100 software package that could be sold to just 1% of the installed base yields sales of over \$50 million.

Over time, the market has demanded increasingly sophisticated software. Each successive enhancement in processor speed has been consumed by software that is more complex, even if only in creating a more user-friendly interface. In the past, computer time was expensive relative to labor costs. That situation is now reversed, and spending more for a more user-friendly computer can often be easily justified in order to enhance the productivity of the vastly more expensive human being.

Whereas the general office community reaped the benefits (and endured the pain) of the rise of the personal computer, the scientific and technical market was forced to wait. Chemists used spreadsheets and suffered through word processors that did not know what a chemical structure was. However, there is now a good variety of personal productivity tools that have been designed specifically for the chemist.

Word processors that deal readily with chemical equations and mathematical formulas are now commonplace. There are several molecular-structure drawing tools, the output of which can be imported into popular word processing packages. There are numerous packages for displaying, manipulating, and analyzing scientific data, as well as many that can be used for data acquisition and laboratory instrument control. Many chemical journals regularly review new scientific software packages.

As is the case elsewhere in the industry, there are no constants. The package currently judged superior is often eclipsed by a newer version of a competing package. The only trend that remains constant is that software continues to become more complex while efforts to make computers easier to use increase. Although the benefits first accrue to the general business community, increasing interapplication cooperation, where the output of one package can be easily integrated into another package or where one package maintains live links to the data in another package, can be expected. Changing data in one package will see those changes automatically propagated to all other applications that are using that same data.

The networking of personal computers has continued and has enabled the sharing of expensive hard-copy output devices and other peripherals and has provided the significant benefit of facilitating the transfer of data between computers. Joint authorship of technical articles is a quick and easy process with networked computers. The interapplication communication that today links applications within a single computer will transparently link applications on networked computers. Voice-annotated documents are already available.

The great majority of desktop computing is performed in the DOS or Macintosh operating environments. Microsoft Windows has been an immensely popular addition to DOS. IBM is currently aggressively marketing OS/2 as the preferred desktop computing environment for the future. Microsoft has strongly supported Windows. Apple has recently formed an alliance with IBM to jointly develop object-oriented methodologies for future desktop computing. It is beyond the scope of this article to speculate on the probability of success of these initiatives, but it can be said that the industry is entering a potentially confusing and rapidly changing era.

Supercomputers

It is difficult to define a supercomputer. Today's supercomputer becomes the minimum expectation for tomorrow's computers. A ten-year-old

supercomputer is most likely a museum piece. Although there may be no universally accepted definition of a supercomputer, there are some characteristics that all supercomputers have. A supercomputer is expensive. Its cost has stayed relatively constant over the years, typically between \$1 and \$30 million. Performance is the primary goal of the designers of a supercomputer; cost is a secondary consideration at best. Supercomputers utilize the fastest electronic components available, connected in ways designed to minimize transmission delays. With a large number of electronic components driven at very fast rates, the typical supercomputer generates enormous amounts of heat. Most require extensive power-supply and liquid-cooling support systems.

Supercomputers are found in many government research laboratories, intelligence agencies, universities, and a small number of industrial companies. In the United States, the National Science Foundation (NSF) has provided supercomputers to several prominent universities for both academic and industrial users. These centers provide state-of-the-art, supercomputer-tuned applications for a wide variety of disciplines, together with staffs who are very knowledgeable in optimization for supercomputer performance.

A common acronym is MFLOPS, millions of floating-point operations per second. Because most scientific computations are limited by the speed at which floating point operations can be performed, this is a common measure of peak computing speed. Supercomputers of 1991 offered peak speeds of 1000 MFLOPS (1 GFLOP) and higher.

Vector Computers. Most computers considered supercomputers are vector-architecture computers. The concept of vector architecture has been a source of much confusion.

Most computers use pipelining to maximize performance. Pipelining is analogous to an automobile assembly line, with the finished product being the result of many sequential but otherwise independent operations on the object being produced. It is inefficient to have only one car on an assembly line at one time because stages which had finished their operations on the car would then sit idle until the last stage had completed operations before they could begin work again. In computing, many operations can be decomposed into suboperations that can be performed sequentially but independently. The pipelining in the Amdahl 470V 6 supercomputer has been described as follows (1975):

A high throughput of instructions was achieved by pipelining the processing of instructions. The execution of instructions was divided into 12 suboperations that used 10 different circuits. When flowing smoothly, a new instruction could be taken every two clock periods (or 64 nanoseconds) and therefore up to six instructions were simultaneously in different phases of execution, and could be said to be in parallel execution (1).

Clearly the key to maximum performance was being able to keep the pipeline full. This observation still holds true for today's computers.

There are many reasons that it might be difficult to keep the pipelines full. The most obvious is a data dependency, where a previously initiated computation must pass through all stages of the pipeline before the next operation can be commenced.

$$X = 2 * Y$$

$$Z = 3 * Y$$

$$W = Z * 2$$

In this example, the evaluation of *X* and *Z* can be overlapped within the multiplication pipeline, but work cannot begin on the evaluation of *W* until the result of the computation of *Z* is known. The pipeline must empty, and the result, *Z*, must be retrieved before the pipeline can be refilled for the evaluation of *W*.

A pipelined floating-point multiply unit might accomplish a floating-point multiply by performing four independent suboperations, labeled a, b, c, and d, on the operands. The suboperations can be envisioned as the four workers on a four-person assembly line. The floating-point multiply pipeline could accept a new set of operands every clock cycle. The pipeline occupancy of this code fragment would look like

clock cycle	0	1	2	3	4	5	6	7	8	9
computation of <i>X</i>	a	b	c	d						
computation of <i>Z</i>		a	b	c	d					
computation of <i>W</i>			a	b	c	d				

In other words, *X* is being computed during cycles 0–3, *Z* is being computed during cycles 1–4, and *W* is being computed during cycles 5–8. The code fragment is complete after nine clock cycles.

A modern optimizing compiler would recognize that *X* and *Z* are independent and would probably reorder the code to be

$$Z = 3 * Y$$

$$X = 2 * Y$$

$$W = Z * F$$

Now the computation of *W* can begin on the fifth clock cycle, rather than on the sixth.

	0	1	2	3	4	5	6	7	8
<i>Z</i>	a	b	c	d					
<i>X</i>		a	b	c	d				
<i>W</i>			a	b	c	d			

The code fragment is now finished after the eighth clock cycle. Note that there are still three clock cycles during which there are idle stages in the multiplication pipeline. The compiler would look for other statements in the code that could be overlapped with those already in process.

The best way to ensure that the pipelines are full is to gather together a complete series of inputs before starting the pipeline. These are the vector registers of a supercomputer. Vector supercomputers typically contain at least eight vector registers, each holding 64 or more floating-point numbers.

Consider the following FORTRAN code fragment:

```
DO 100 I = 1, 1000
  X(I) = 2.0 * Y(I)
100 CONTINUE
```


On a vector computer having vector registers that hold 64 floating-point numbers, this loop would be processed 64 elements at a time. The first 64 elements of Y would be fetched from memory and stored in a vector register. Each iteration of the loop is independent of the previous iteration, so this loop can be fully pipelined, with successive iterations started every clock cycle. Once the pipeline is filled, the result, X , will be produced one element per clock cycle and will be stored in another vector register. The results in the vector register will then be stored back into main memory or used as input to a subsequent vector operation.

A significant amount of machine overhead is involved in setting up a vector operation, with maximum benefit accruing if there are a full 64 elements to compute. For short loops, typically eight elements or fewer, vector operations are often slower than doing the computations one at a time in nonvector mode.

Vectorization becomes more difficult when there are loop dependencies, or when one iteration of the loop depends on another. Consider the following:

```
DO 100 I =
X(I) = X(I + N) + 3.4
100 CONTINUE
```

If N is positive, the loop can be vectorized. If $N = 5$, the existing 10th element of X is needed in order to compute the new value of the fifth element. As work begins on the fifth element before the tenth element is changed, the loop can be vectorized.

If, on the other hand, N is negative, then there is a loop dependency, and it may not be possible to fully vectorize the loop. If $N = -1$, the new value of the fifth element of X needs to be known before computation of the new value of the sixth element can begin. It might be several clock ticks before the new value of $X(5)$ becomes available. On the other hand, if N is a larger negative number, for example, $N = -50$, $X(1)$ is needed to evaluate $X(51)$; then it may be possible to vectorize this loop—the new value of $X(1)$ would be known long before the computation of the new $X(51)$ is started. In the absence of knowledge of the possible values for N , most compilers would not produce vector instructions for this loop. If the programmer has knowledge of the possible values of N , explicit directives can be inserted in the code to inform the compiler of assumptions that can be made about the loop.

There are vastly more complex examples of difficult vectorization decisions. A great deal of effort has been devoted to writing vector code, or code that compilers can safely translate into vector instructions. As compilers become more sophisticated, their ability to recognize vectorization opportunities increases. The vendors of vector computers often claim that vectorization is automatic and that end users need not be concerned with it. This claim is usually true only if the end user is similarly unconcerned with achieving maximum performance. More often than not, codes that have not been written for vector architecture machines must undergo substantial restructuring in order to achieve significant performance enhancements on vector-architecture machines.

Banked Memory. Another characteristic of many vector supercomputers is banked memory. The main memory is usually divided into a small number of electronically separate banks. A given memory bank can absorb or supply operands at a much slower rate than the rate at which the central processing unit (CPU) can produce or use data. If the data can be spread across multiple memory banks, the effective memory bandwidth, or rate at which memory can absorb or supply data, is increased. For example, if a single memory bank can supply one operand every 16 clock cycles, then 16 memory banks would enable the entire memory subsystem to deliver one operand per clock cycle, assuming that the data come sequentially from different memory banks.

The memory subsystem on most supercomputers is organized to support maximum performance on loops of stride one, or when the elements of an array are accessed sequentially with no gaps. In general, the stride is defined by

```
DO 100 I = START, STOP, STRIDE
X(I) = .....
100 CONTINUE
```

The most common loop has stride = 1. Typically $X(1)$ would be stored in memory bank 1, $X(2)$ in memory bank 2, $X(16)$ in memory bank 16, and $X(17)$ in memory bank 1. In the loop in the example, if stride = 1, then the elements of X can be delivered to the CPU at the maximum rate, one per clock cycle.

If, on the other hand, stride = 2, then the system memory may limit the speed of the calculation. During the first clock cycle, a request is sent to bank 1 for $X(1)$. During the second clock cycle, a request is sent to bank 3 for $X(3)$. During the 8th clock cycle, a request is sent to bank 15 for $X(15)$. On the 9th clock cycle, when $X(17)$ should be fetched from memory bank 1, that memory bank is still processing the previous request, because $X(1)$ initiated during the first clock cycle. The system must wait until that request is completed before initiating a new request to memory bank 1. This access pattern uses only half the elements of X . More importantly, it uses only half the memory banks available. Consequently, effective memory bandwidth is halved. The most pathological case comes with stride = 16, when all references come from a single memory bank; effective memory bandwidth is one-sixteenth theoretical maximum.

Other Performance Considerations. Even if a program allows main memory to supply operands at peak rate, it may not be fast enough to keep the CPU operating at its peak rate. Consider the general SAXPY

```
DO 100 I
Z(I) = ALPHA * X(I) + Y(I)
```

The term SAXPY has arisen as a mnemonic for scalar alpha X plus Y (2). This loop requires two operands and produces one result for each iteration of the loop. In 64 bit, or 8 byte, precision (8 bytes per floating-point number), this is a total of 24 bytes of data being consumed or produced per iteration. If the memory bandwidth were 240 megabytes per second, the memory subsystem could maintain this loop at 10 million iterations per second. Each loop iteration represents two floating-point operations, a multiplication and an addition; thus running at 10 million iterations per second is only 20 MFLOPS, a small fraction of the peak performance of supercomputers. Most supercomputers have memory subsystems with much higher bandwidths, sometimes with separate pathways for read and write operations. Nevertheless, careful analysis of memory subsystem usage can be an important ingredient of any code optimization. In the preceding example, the system should look for subsequent operations to be performed on $Z(I)$ while it is still near the CPU, before it is returned to main memory. The problem of optimizing CPU performance is not specific to supercomputers. However, given the enormous cost, a great deal more effort is devoted to optimizing codes on supercomputers than on other machines.

Because of the relative slowness of main memory (compared with the CPU), most computers have a much smaller, but much faster cache memory subsystem that augments main memory. The size of the cache memory and the extent to which a program can utilize the cache can be critical determinants of performance. Again, there are some common optimization techniques designed to maximize cache utilization.

FORTTRAN stores doubly dimensioned arrays in memory as:

$$X(1,1) \ X(2,1) \ X(3,1) \ ... \ X(n,1) \ X(1,2) \ X(2,2) \ ... \ X(n,2) \ X(1,3) \ ...$$

This has important consequences for programs that use such arrays. Consider the two code fragments:

Version 1

REAL X(M,M)
DO 200 I = 1, M
DO 100 J = 1, M
X(I, J) = 0.0
100 CONTINUE
200 CONTINUE

Version 2

REAL X(M,M)
DO 200 I = 1, M
DO 100 J = 1, M
X(I, J) = 0.0
100 CONTINUE
200 CONTINUE

The programs are functionally identical and produce the same result. Version 1 accesses $X(1,1), X(1,2), X(1,3) \dots$, whereas version 2 accesses $X(1,1), X(2,1), X(3,1) \dots$. Version 2 executes much faster on most computers, because it can exploit cache memory much more effectively than can the first version. Version 2 accesses the elements of X in the same order in which they are stored in memory, or in cache. On first reference to array X , a portion of the cache is filled with contiguous elements of X . Whereas version 2 of the program accesses all the elements of X now in cache, version 1 only accesses one element before making a reference to $X(1,2)$, which, being from a potentially distant memory location, is unlikely to be already in cache; a portion of array X from $X(1,2)$ is loaded into cache, perhaps displacing the portion near $X(1,1)$. Version 1 is likely to run no faster than the speed of memory. Again, this phenomenon is characteristic of virtually all computers having cache memory and is not limited to supercomputers. Vastly more complex examples of optimization for maximum cache utilization exist. Because supercomputers are typically much more expensive than the labor of the people who use them, these painstaking optimizations can pay big dividends.

The hierarchy of disk, main memory, and cache, each one faster than the one before, is a general one. On the top of this pyramid are registers. Supercomputers typically contain a small number of ultrafast registers within the CPU. The registers hold scalar variables, such as loop counters, or accumulator variables. At all stages of the hierarchy, performance tuning involves maximizing the use of the faster components. The compiler usually decides which variables should be kept in registers.

Performance. The most commonly cited performance measure in the scientific computing world is the LINPACK benchmark (3). The benchmark involves diagonalizing a 100×100 double precision matrix. The 100×100 problem is moderately vectorizable in that good speedups over scalar execution can be achieved, but it generally does not permit vector computers to perform near peak performance. Table 1 contains an extract from the August 1991 LINPACK report. LINPACK data is reported as MFLOPS, so larger numbers are better. Because LINPACK is such a well-known benchmark, a great deal of effort is devoted to optimizing its performance on many computers. Successful efforts tuning LINPACK may or may not correlate with increased performance for other programs.

Table 1. Sample LINPACK MFLOPS Ratings,^a Aug. 1991

Computer	MFLOPS ^b	Computer	MFLOPS ^b
Cray Y-MP 16 prototype	403	Alliant FX 2800-200 ^c	10
NEC SX-3 14	314	MIPS RC6280	10
Cray Y-MP 832 ^d	275	Stardent 3010	10
Fujitsu VP2600 10	249	IBM RS 6000 model 320	9
Cray Y-MP 832	161	SCS-40	8
Cray Y-MP 416	121	Convex C-130	7.2
Hitachi S-820 80	107	DEC VAX 6000 (vector)	7.0
Cray 2S 4-128	107	IBM 3090 180	6.8
ETA 10-G	93	Alliant FX 2800-200	6.4
Cray Y-MP 832	90	Silicon Graphics 4D 420	6.0
IBM ES 9000 model 900 VF	60	FPS 264	5.9
Convex 3810	44	MIPS RC3360	4.5
IBM ES 9000 model 900	38	SUN SparcStation II	3.8
CDC Cyber 2000V	32	DecStation 5000 200	3.7
Alliant FX 2800-200 ^e	29	Gould NP1	3.1
Cray 1S	27	IBM 370	2.5
IBM RS 6000 model 550	27	SUN SparcStation I +	1.6
DEC VAX model 9000	22	DEC VAX 6000 model 410	1.2
FPS model 522	20	SUN 4 260	1.1
Amdahl 1400	19	DEC VAX 8650	0.70
Silicon Graphics 4D 480 ^d	18	SUN 3 260 (FPA)	0.47
Convex C-210	17	Apple Macintosh IIfx	0.41
Cydrome CYDRA 5	14	VAX 11 785 (FPA)	0.20
Cray 1S (1983)	12	Compaq 386 20 w 387	0.16
Multiflow Trace 7 300	11	DEC microVAX II	0.13

^a Ref. 3.

^b All data are for single processor execution except where noted. Higher numbers indicate better performance. As compilers change, these ratings may change.

^c Two processors.

^d Eight processors.

^e Twelve processors.

Performance achieved on single processor vector computers is governed by Amdahl's law (4). Once started, vector operations can be performed much faster than single arithmetic operations. As an example, consider a machine in which this speed ratio is 10:1, a very low ratio. Assume further that the program runs in 100 s without using vector instructions. If 10% of that program can be run in vector mode, and thereby speeded up by a factor of 10, execution time drops to 90 + 1 = 91 s. If, on the other hand, 90% of the program can be run in vector mode, execution time is now 10 + 9 = 19 s. This situation is illustrated in Figure 1.

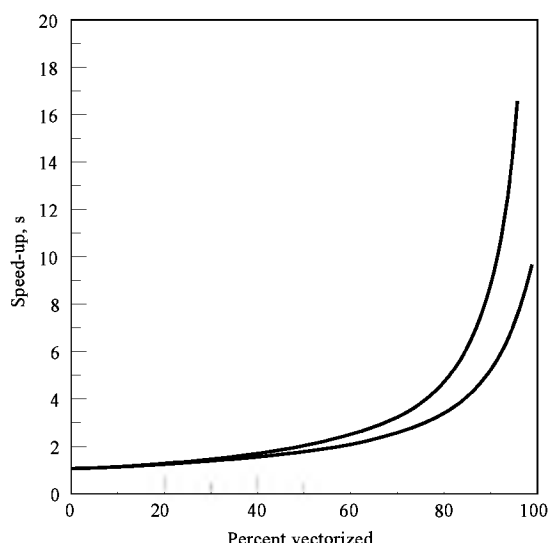


Fig. 1. Amdahl's law. Speedup as a function of the percentage of the program that can be vectorized. Lower curve vector-scalar speedup = 10; upper curve vector-scalar speedup = 100.

Thus, only when substantial parts of the program can be run in full vector mode can significant overall speed improvements be achieved, and regardless of how efficiently the part of the program that can be run in vector mode is run, the nonvector part will still limit the execution time. Most real world, as opposed to synthetic benchmark, programs contain significant nonvectorizable portions. For this reason, it is now widely accepted that traditional vector supercomputers will not be the architecture that first delivers usable TERAFL0P (1000 GFLOP) performance.

Peak Performance. Every computer has a theoretical peak speed, the speed that would be achieved if all the pipelined functional units of the machine could be kept fully supplied with operands. It is a speed that is guaranteed never to be exceeded. These peak processing speeds make for interesting discussion and speculation; however, few programs allow vector computers to run at anything approaching peak speed. Recent LINPACK reports contain a section that describes the performance of a program that does allow vector computers to perform near peak speeds.

Supercomputers from vendors such as Cray, NEC, and Fujitsu typically consist of between one and eight processors in a shared memory architecture. Peak vector speeds of over 1 GFLOP (1000 MFLOPS) per processor are now available. Main memories of 1 gigabyte (1000 megabytes) and more are also available. If multiple processors can be tied together to simultaneously work on one problem, substantially greater peak speeds are available. This situation will be further examined in the section on parallel computers.

Chemistry.

Many computational chemistry programs have been adapted for vector processing, often with good gains in speed. The work on adapting the programs has been done both by the original developers and by staffs at the NSF centers. Many of the optimizations are related to the use of optimized matrix manipulation routines or to recasting data structures into forms that are amenable to vectorization.

Ab initio quantum chemistry programs (5,6) can easily be made to occupy any computer fully, and so are good candidates for a supercomputer. These programs treat individual atomic orbitals as combinations of Gaussian orbitals. Generally, the more Gaussian orbitals used in the basis set, the more accurate the results. As the problem scales as the fourth power of the number of electrons (or the sixth power when using theories that account for electron correlation), it is easy to cast a problem that requires the generation and use of several gigabytes of data. Early versions of the programs computed these two electron integrals, stored them on disk, then fetched them back into memory as they were needed for subsequent calculations. Over the years, the speed of disk drives has increased slowly, whereas CPU and compiler technologies have made much more rapid progress. It is now often more efficient to recalculate the two electron integrals rather than to wait for previously computed values to be retrieved from a disk (7).

Supercomputer Costs and Benefits. Given the great expenses associated with supercomputers, not only the initial purchase and facilities cost but also the large support staffs typically required, they are most often shared by large numbers of people, with the undesirable effect of providing to each user only a small fraction of the computer's performance. Sometimes in these situations no user gets supercomputer-class throughput. On widely shared supercomputers, large jobs are often run overnight only in order to maximize the availability of resources for daytime users. There is a great deal of debate about the ultimate value of widely shared supercomputers. Well-funded, grand-challenge, or mission-critical applications can often justify their own dedicated supercomputers; however, such projects seem to be more the exception than the rule.

Minicomputers

In the late 1970s and early 1980s, the typical chemistry department used a central time-shared minicomputer for virtually all its computing. These air-cooled computers required neither expensive liquid cooling nor large support staffs. They were priced within the range of typical departmental budgets and could be run according to the needs of the individuals in the department. As many as 30 users would share such a computer, often providing each user much less throughput than that provided by a personal computer. However, the cost and flexibility of these systems overwhelmed other concerns, and the minicomputer became the standard for departmental scientific computing during the 1980s, when computers still cost much more than the labor of the people using them. Painstaking efforts were undertaken to optimize computational chemistry programs for the minicomputer architectures. With computations taking days or more to run, even small percentage gains could translate to time savings of hours.

The Digital VAX rose to prominence as a departmental minicomputer and became a virtual standard in the world of chemistry. The VAX offered a user-friendly flexible environment, together with what was then considered good computational throughput. Much computational chemistry methodology was developed on the VAX.

Chemistry. The widespread availability of minicomputers and the advent of robust programs led to greatly enhanced use of *ab initio* and semi-empirical quantum chemistry techniques (8). Whereas it might once have been necessary to negotiate with computer center staff for computer time and to consult with a quantum chemistry expert in order to perform detailed structure calculations, minicomputers and easy to use quantum chemistry programs made these computations much more readily available to researchers. This trend of greater availability, enhanced ease of use, lower cost, and more power seems to be a general phenomenon, although in the case of quantum chemistry the seeming ease of use has sometimes hidden the limitations inherent to the techniques. This trend poses interesting challenges to all disciplines in which computational results can be subject to interpretation, but are easy and cheap to perform.

As the minicomputer rose to prominence so did the more widespread use of molecular mechanics as a computational technique. Whereas the quantum chemical programs deal with molecules as nuclei and electrons, the molecular mechanics paradigm treats each atom as a classical ball of a certain mass. The bonds connecting the balls are treated as classical, generally harmonic springs, and bond angles are described by similar classical terms. Through space (London), rather than through bond, interactions are typically described by Lennard-Jones potential functions.

The strength of molecular mechanics is that by treating molecules as classical objects, fully described by Newton's equations of motion, quite large systems can be modeled. Computations involving enzymes with thousands of atoms are done routinely. As computational capabilities have advanced, so

have the size and complexity of the systems modeled with molecular mechanics. Unfortunately, the execution time can scale as the square of the number of particles, but the method does hold interesting possibilities for parallel processing, which will be discussed later.

Much of the early work in molecular mechanics has been commercialized, with development continuing both in universities and by software vendors. As commercial products, these programs have been "cleaned up" with the addition of user-friendly interfaces, proper documentation, and dedicated support staff. With commercialization has also come marketing, with competitive pressures driving vendors to aggressive marketing techniques. Also, algorithmic developments that once were freely shared among colleagues are often now considered trade secrets. Chemists have certainly benefited from some aspects of the commercialization of molecular mechanics tools; however, its long-term effects are not at all clear. Again, the minicomputer was the initial enabler of what has become a pervasive technique.

The preeminent offerings in this crowded market include MacroModel (Columbia University, New York), Insight (Biosym Technologies, California), Sibyl (Tripos, Missouri), ChemX (Chemical Design, UK), BioGraf (Molecular Simulations, California), Charmm Quanta (Polygen Corp., Massachusetts), PC Model (Serena Software, Indiana), ChemLab (ChemLab Inc., Illinois), and a large number of personal computer-based packages.

A truism of computational chemistry is that chemists will always want to model ever larger systems, or smaller systems, at ever more accurate levels of approximation. The total running time of jobs has, in general, not lowered dramatically. Computational chemists still perform calculations that take several days to complete. However, today the molecules can be much larger and the quality of the calculations better.

Work Stations

The mid-1980s was a turning point for minicomputers as microprocessor-based UNIX work stations began to appear. Development of the UNIX operating system began at Bell Labs in 1969. UNIX was an operating environment built for programmers by programmers. During the mid-1970s it was freely distributed to universities, many of which made significant enhancements to it. SUN Microsystems marketed a series of microprocessor-based work stations, which ran an enhanced version of UNIX from Berkeley, California. These systems offered large, high resolution, multiwindow monitors, together with computational speed that rivaled or exceeded that available on the minicomputers of the time. There began a transition that later became a stampede.

Vendors such as SUN and MIPS introduced lines of computers based on RISC (reduced instruction set computer) chips. These computers offered significant performance advantages over the CISC (complex instruction set computer) minicomputers, at least for CPU-bound work. Although there are still active debates about what RISC and what CISC are, the essence of RISC is simplicity.

The philosophy of RISC is that the CPU performs a very small number of very simple operations. Whereas a CISC-based computer might have an instruction that fetches a number from memory and updates a counter, a RISC system implements such an operation with multiple, but simple, instructions. By keeping the CPU simple, it can be more readily scaled up to ever greater speeds. The idea is that, although it might execute many more instructions than a CISC machine, it can perform its simple instructions so much faster that it gets more work done in a given time period.

The RISC versus CISC conundrum has led to the much abused and ultimately extremely confusing term MIPS (millions of instructions per second). Measures of performance that can be more directly related to a computer's ability to perform useful work should always be preferred over machine MIPS. The throughput of a computer is a function of the number of instructions to be executed, the average number of instructions that can be executed per clock cycle, and the time per clock cycle.

An additional advantage of the RISC microprocessor computers is that their implementors laid plans with a view of semiconductor and compiler technology of the 1990s. Most of the earlier CISC systems were defined with a view of what the technology of the 1980s would bring. RISC-microprocessor-based computers hold an increasing performance advantage over CISC-based systems.

Table 2 shows timings for running version 5.0 of MOPAC (9), a semi-empirical quantum chemistry program, on a series of computers commonly used in chemistry environments. Because the table measures elapsed time, lower numbers are better. Although MOPAC 5.0 is an important program for many chemists, it is not a generally accepted industry standard performance measure. Caution should be exercised interpreting benchmarks, especially when they are run by others. Comparing the LINPACK and MOPAC benchmark data is illustrative.

Table 2. Performance According to MOPAC 5.0 Benchmark Data^a

Computer	Cycles ^b	Elapsed time ^c
SUN 3 280 ffpA	29	39:00
SUN 4 260	29	24:18
Sparcstation 1	29	19:26
DecStation 3100	29	12:50
MIPS RC3240	29	7:08
DecStation 5000 200	29	6:18
SparcStation II	29	6:18
Silicon Graphics 4D 35	29	4:50
IBM RS 6000 model 320	29	2:40
IBM RS 6000 model 530	29	2:02
Hewlett Packard model 720	29	1:59
IBM RS 6000 model 550	30	1:19
VAX 8600	30	36:04
IBM 3090-200	30	10:08

^a Geometry optimization of a small molecule with version 5.0 of MOPAC (9).

^b Number of geometry optimization cycles.

^c Computation times, mins.

In Table 2, the computers in the MOPAC benchmark are grouped as they are to reflect differing floating-point machine representations. All these computers use 64 bits to represent a double precision floating-point number. Different behavior arises from the slightly different ways in which the bits are allocated.

Geometry optimizations in MOPAC are iterative in nature, and the program may go through a different number of cycles to achieve the same convergence criterion under different floating-point representations. In this case, the program goes through an extra cycle on both the VAX and IBM mainframe compared with most IEEE machines. Even factoring in the extra cycle it is obvious that the IEEE machines run MOPAC exceptionally well. Whereas LINPACK ranks an IBM 3090 as being significantly faster than a DecStation 5000 200, the MOPAC 5.0 benchmark shows just the opposite. The importance of well-characterized benchmarks should be clear; what is best for one person may be suboptimal for another.

The widespread confusion about MIPS and MFLOPS, together with the existence of programs that exhibit widely differing behavior on different computers, led to the formation of the SPEC group, Systems Performance Evaluation Cooperative. This group maintains the SPEC benchmark, a suite of programs having differing characteristics. Some are limited by integer performance, others by floating-point performance; some benefit from vectorization, others do not. By reporting a performance metric that is an average over several programs, it is anticipated that the SPEC rating will be more generally predictive than are benchmarks based on a single program. Of course, for those who use primarily floating-point intensive programs, the integer intensive benchmarks in the SPEC benchmark may not be of great interest; however, the purpose of the SPEC suite is to provide a general purpose rating. Note, however, that compilation is an integer intensive operation. The SPECmark rating of several well-known RISC work stations is included in Table 3.

Table 3. SPECmark Ratings^a for Popular RISC Computers

Computer	SPEC int	SPEC fp	Rating ^b
Hewlett Packard model 730	51	100	76
IBM RS 6000 model 550	34	119	72
Hewlett Packard model 720	39	78	59
MIPS RC6280			42
IBM RS 6000 model 320	16	53	32
Silicon Graphics 4D 35			31
SUN SparcStation II	20	21	21
DEC Decstation 5000 200	18	20	19
DEC Decstation 3100			10
SUN SparcStation I			8

^a Where available, the SPEC integer (int) rating and the SPEC floating-point (fp) rating are reported separately.

^b Ratings are normalized to a VAX 11 780. As compilers change, SPEC mark ratings may change.

Many of the faster work stations can provide throughput similar to that observed on a crowded, shared supercomputer, especially for codes that do not benefit greatly from vectorization. The availability of such machines for less than \$50,000 (much less for academic users) has once again changed concepts of what is computationally feasible. Many more people can perform computations that a few years ago were the sole domain of those with access to large-scale computing facilities, and this trend is expected to continue.

RISC work stations have been doubling in performance every two years since the mid-1980s. This growth is much more aggressive than that observed in other technologies (supercomputers, mainframes, minicomputers, etc). This trend must slow down eventually, as limits imposed by constraints such as the speed of light are approached. Computer vendors, however, are confidently predicting that the trend will continue unabated at least until the mid-1990s.

As CPU performance increases, the gap between CPU and disk and memory speeds will continue to widen. As limits of technology are approached, other techniques will be needed to gain performance advantages; more functional units, multiple processors, and so on. These approaches are discussed in the sections on minisupercomputers and parallel processing.

Technical RISC work stations have revolutionized computational science in five years.

Visualization

With the ever-increasing ability of more and more people to do more and more computation than ever before, the sheer volume of computational data produced can be overwhelming. As computational availability has increased, the rate-limiting step to many computational studies has become interpretation of the results, rather than waiting for the computations to complete. In any discipline, interpreting a 10-cm-thick computer printout is a daunting task. Although humans are not efficient when reading numbers sequentially, they are extremely efficient at visual interpretation because visual information is processed in a parallel fashion. Thus, if a mass of data can be presented in a meaningful visual form, the analysis of computational science can be greatly facilitated. Visual presentation not only speeds up interpretation, but also makes readily discernible insights into the results that are virtually unobtainable from a pile of paper. Visualization aids in the initial interpretation of the data and also in its communication to others.

Early visualization systems involved a special purpose graphics terminal attached to a minicomputer or mainframe time-shared host. Just as RISC work stations eclipsed mainframe and minicomputer systems for CPU-bound work, so, too, did visualization migrate to smaller systems. The additional advantage of the work station is that the graphics subsystem can be tightly integrated with the CPU, leading to a mutually supportive combination of computing power and simultaneous visualization of the results. Just as RISC CPU power is increasing dramatically, the market is also driving a similar race for graphics performance, with ever greater drawing and rendering speeds being required. Being able to rotate a three-dimensional depiction of a molecule, perhaps colored by atomic charge, or a depiction of the electrostatic field around a molecule, can lead to great insights into detailed molecular behavior.

Two successful and widespread applications of visualization techniques in the field of chemistry are the visualization of molecular orbitals and the visualization of molecules in molecular mechanics studies.

It is now possible to "see" the spatial nature of molecular orbitals (10). This information has always been available in the voluminous output from quantum mechanics programs, but it can be discerned much more rapidly when presented in visual form. Chemical reactivity is often governed by the nature of the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO). Spectroscopic phenomena usually depend on the HOMO and higher energy unoccupied states, all of which can be displayed and examined in detail.

All the molecular mechanics programs mentioned in connection with mini-computers now run on RISC work stations with integrated graphics systems. Three-dimensional depictions of molecules can be rotated, docked with other molecules, and otherwise manipulated interactively. Important breakthroughs in understanding drug action have come about from the ability to visualize substrate molecules interacting with receptor sites. For example, there is a lock-and-key mechanism in which the substrate or inhibitor molecule must fill a well-defined cavity within the protein and interact with specific functional groups within that cavity. Molecular visualization greatly facilitates the discovery and communication of such ideas.

The challenges for visualization are at least twofold. Faster graphics hardware will be required to display and manipulate more complex data displays. More importantly, the human effort required to develop visualization systems must be reduced. It is the realm of the expert programmer to implement a usable visualization system. General purpose tools that allow the nonexpert to import data in different formats into robust visualization systems are just beginning to appear.

The evolution (revolution) of capabilities has dramatically changed the way in which chemists work, and, again, many more people can now perform and analyze many more computations than ever before, at ever diminishing costs.

Minisupercomputers

The so-called minisupercomputers that emerged and then declined during the 1980s were for many years some of the most interesting computers in terms of architecture. Many of the more successful aspects of their designs are expected to be incorporated into general computer design practice. The minisupercomputers were typically minicomputer-sized systems that offered performance levels of one-quarter to one-third of that available on the supercomputers of the time. As their prices were close to minicomputer prices, they were an attractive alternative for many who needed supercomputer performance but did not have supercomputer budgets. A dedicated minisupercomputer might provide more throughput than a crowded, shared supercomputer.

For all the excitement and enthusiasm of the computer architects, these computers did not meet with great success in the marketplace, and few companies remain as viable entities. One of the primary reasons for their demise seems to have been the simultaneous rise of the RISC work stations, which killed off numerous other architectural initiatives, hence the term *killer micros* (11).

Having a marketplace crowded with differing computer architectures, each perhaps requiring different strategies for achieving maximum performance, mini-supercomputer vendors laid siege to application developers to get key applications ported to their systems. Given the potentially great costs and possibly small returns from such endeavors, these overtures were often in vain, leading to a vicious cycle of a paucity of applications inhibiting

sales, and no applications because of the small installed base. The most successful porting and optimizations usually involved a partnership between the application developer and the computer vendor, but the costs were often high.

The most commercially successful of these systems has been the Convex series of computers. Ironically, these are traditional vector machines, with one to four processors and shared memory. Their Craylike characteristics were always a strong selling point. Interestingly, SCS, which marketed a minisupercomputer that was fully binary compatible with Cray, went out of business. Marketing appears to have played as much a role here as the inherent merits of the underlying architecture.

Multiflow. One of the most interesting and innovative entrants into the minisupercomputer field was Multiflow (12). The Multiflow Trace series was a VLIW, very long instruction word, architecture computer, with a great deal of fine-grained parallelism in the architecture. Each CPU contained seven independent functional units: conditional branch, two integer operations, two memory operations, and two floating-point operations. With so many functional units potentially simultaneously active during each clock cycle, a very long instruction word was needed in order to specify just what each functional unit would do on each clock cycle. Each CPU required a 256-bit instruction word, and up to four CPUs could be combined into one machine, leading to a 1024-bit instruction word. The system used banked memory, but unlike more expensive supercomputer systems that used hardware to control memory-access conflicts, the Multiflow controlled conflicts at the compiler level. The compiler became vastly more complex in order to avoid costly and complex hardware components.

The Multiflow achieved very high rates of performance using electronic components, which were relatively slow for the time. Its performance was achieved because of its ability to keep the multiple functional units busy. The compiler would aggressively rearrange code in order to achieve this goal. Once the operands for a particular instruction were available (computed or fetched from memory), that instruction could be scheduled, regardless of where it first appeared in the original code. Clearly the correctness of this approach requires a careful dependency analysis of the code.

Whereas conditional branches often destroy parallelism in more traditional computers (13), the Multiflow employed several strategies for dealing with branches. With the code fragment below, most traditional computers would need to wait until the comparison with A was complete before beginning evaluation of $D = E + F$.

$$A = B + C$$

$$IF (A.GT.1.0D + 03) D = E + F$$

$$I = I + 1$$

Because the Multiflow had multiple functional units, it would simultaneously perform both $A = B + C$ and $D = E + F$, storing both results in registers. If the result of the comparison turned out to be false, then it would not store the result D back into memory, and the calculation would be discarded. Because the two operations were done in parallel, this method took no extra time. The integer operation, $I = I + 1$, as well as four other operations, could also be performed simultaneously.

Perhaps the most unusual aspect of the compiler is that it would make compile time decisions about the most likely outcome of a comparison operation and generate code to follow that most likely path. The compiler would insert compensation code to undo calculations already done if the branch were other than predicted. In addition, it was possible to execute a program and monitor the results of branch decisions. Armed with more accurate information about the likely outcome of conditional branches, the compiler could then generate better code, because its "guesses" would be correct more often.

Loops that could not be vectorized on conventional vector computers often performed very well under the Multiflow architecture; and, unlike vector machines, for which a person could spend a great deal of time optimizing programs, substantially less could be done on the Multiflow, as most of the work fell to the compiler anyway. All the usual optimizations for memory utilization and cache usage also applied to the Microflow. There were, of course, programs for which the compiler could not make good use of the multiple functional units, and the computer would run at the speed of just one or two individually quite slow functional units.

Others. Other architecturally interesting, but commercially unsuccessful, computers were developed by Cydrome (14), Astronautics (15), Gould, Vitesse, and others. All were creations of the ready availability of capital for computer start-up ventures in the early 1980s, but fell victim to the resulting crowded marketplace, poor marketing, and the killer micros; why work to tune a 3-MFLOPS (scalar mode) vector minisupercomputer to 10 MFLOPS performance when a killer Micro would run the original code at 10 MFLOPS with no changes at all? The minisupercomputers have retreated to the niches of highly vectorizable codes, very large memory requirements, or high input/output (I/O) requirements. As the killer micros infringe on even these areas (16), the future viability of the minisupercomputers remains an open question.

Parallel Processing

The vast majority of the speed increases previously described in the uniprocessor world are, in fact, parallel in nature: multiple functional units, multiple pathways to and from memory, pipelined operations, and so on. These have always been within the context of a single, tightly integrated CPU unit that executed a single sequential stream of instructions. With the speeds of such uniprocessors approaching insurmountable physical limitations, the next great leap in computational throughput can only come from parallel processing, that is, having more than one processor cooperatively working on the same problem at the same time.

Many of the problems inherent in parallel processing are illustrated by the following anecdote.

A nobleman of the Middle Ages derived great pleasure from watching the mowing of his estate. Each month he would summon a mower from the nearby town and would then sit and watch the mowing. This typically took four hours. The nobleman reasoned that if he were to get two mowers, the mowing would be completed in two hours. This he verified the next month. He then reasoned that if he obtained the services of every mower in the country, the mowing would be completed within a matter of three seconds.

The deficiencies in this reasoning are obvious. Similarly, there often comes a point at which adding an extra processor to a problem may decrease throughput rather than increase it; imagine when the area to be mown by each person becomes smaller than a scythe swipe.

The lawn-mowing analogy is also interesting in that up to a certain point there will be a linear speedup as mowers are added. This speedup occurs because the mowers do not interfere with each other's work and have efficient mechanisms for coordinating their efforts. The lawn-mowing problem exhibits very good data parallelism.

Mutual Exclusion (MUTEX). The idea of multiple entities all working on the same piece of work raises this issue of coordination and communication among the individual processes. A well-known example from banking is instructive. Consider two bank tellers simultaneously performing withdrawals from the same bank account. Both read the account balance and determine that the balance is \$100, and so a withdrawal of \$100 is allowed. Both then withdraw \$100 from the account. Clearly this process needs some means by which the actions of the independent processes can be synchronized and coordinated.

The notion of an atomic operation is important for synchronization. An atomic operation is one that is indivisible. Once initiated, it will continue to completion. There are usually a large number of synchronization primitives in a parallel computer, most commonly test and set primitives, or semaphores implemented in hardware (10). A test and set operation tests the current value of a variable and optionally sets a new value, all in one indivisible operation. A semaphore forces the serialization of multiple processes around a critical section, a part of a program that must be executed by only one process at a

time, such as changing the balance of a bank account, for example. When multiple processes request the same semaphore, they are automatically serialized. Of course, in systems with multiple semaphores, deadlock conditions can easily arise. For example, process 1 has control of resource R1 and needs resource R2 to continue; process 2 has control of resource R2 and needs resource R1 to continue. These situations are either avoided by careful design or must be detected and resolved as they occur. This is generally a nontrivial problem that could consume significant resources.

Parallel Languages. The computer science community generally abhors FORTRAN and has been predicting its demise for some time. A relatively new realization is that the scientific community will not be abandoning FORTRAN any time soon and that parallel computing must be fully available within a FORTRAN context. The scientific community has been slow in changing to languages that inherently express parallelism. Nevertheless, there are vigorous ongoing efforts in developing parallel programming environments and languages (17).

Performance Boundaries. Some fundamental laws work against parallel computers. For example, the same program will always run slower on a two-processor parallel computer than on a uniprocessor having double the processor speed, because the parallel computer must spend processing time synchronizing the work of its two processors, a task that the uniprocessor does not need to perform. There is also no guarantee that a program can be broken down into two computationally equal parts.

A recent victim of the killer micros was Evans and Sutherland's parallel computer development effort, halted in 1990. Their architecture combined a small number of approximately 1-MFLOPS processors into semi-independent functional units. Several of these units could, in turn, be combined to form a processor hierarchy, building up to systems that were expected to cost between 1 and 8 million dollars. With the advent of 10-MFLOPS uniprocessor killer micros, such an architecture became irrelevant and the project was halted. The RISC killer micro could deliver the same level of performance as could the combined efforts of 10 of the 1-MFLOPS processors, even with the unlikely assumption that the problem could be perfectly parallelized across 10 processors.

Speedup. The term *good performance* merits discussion. The ideal parallel computer has as many as an infinite number of processors, as much as an infinite amount of zero-latency shared memory, and all interprocessor communications require zero time. If an infinitely parallelizable problem takes T seconds to execute on one of the processors, then it will take T/N seconds to execute on N processors. More commonly this is referred to as an N -fold speedup. Real computers and problems are different, and full N -fold speedup with N processors is impossible. One measure of the efficiency of a parallel computer is how close it comes to achieving N -fold speedup with a given program. This is a strong function of both the program and how well the parallelism inherent to the problem can be mapped onto the parallelism of the computer. In many cases problems may be too large to run on a single processor, making speedup measurements difficult.

The other kind of performance scaling is, for a constant number of processors, how does the running time change as the size of the problem increases? With more grid points, more orbitals, more molecules, and so on? Whereas a uniprocessor may exhibit m^3 behavior, where m is some parameter of the problem, a different, parallel algorithm may exhibit a better-size scaling behavior, growing slower than the serial algorithm. As with many asymptotic effects, this may only be significant for large m , but it is frequently a regime of great interest. If the inherently serial parts of the problem scale unfavorably with increased problem size, those parts may come to dominate performance, but this is rare.

Cache Coherence. Another effect that leads to reduced throughput from computers with multiple independent processors is the issue of cache coherence (18). The performance of many of these machines is critically dependent on the cache utilization of the program. For efficiency, individual processors may store frequently used variables in cache only and not write the value to memory immediately (a so-called write-back cache system). A problem arises when another processor needs to access the current value of that variable. The most current value for that variable may be in another processor's cache memory, rather than in main memory. Multiprocessors with cache memories require explicit mechanisms for ensuring cache coherency among the processors. In many such systems the trend has been to implement a separate bus used just for such interprocessor communication and synchronization operations, rather than taking up bandwidth from the main system memory bus. Nevertheless, a shared bus architecture will always be of decreasing effectiveness as more and more processes are added, because the bandwidth shared among the processors is finite.

The obvious solution to the limitations imposed by shared bus communications is to fully connect each processor to all other processors via dedicated pathways. The problem is that the number of such pathways grows rapidly, $N(N-1)/2$, where N is the number of processors. The inherent costs and complexity of such a system render it an impractical solution for large-scale parallel computing.

Types of Computers. Computers can be classified by Flynn's taxonomy (19). The three important classes are SISD: single instruction, single data; SIMD: single instruction, multiple data; and MIMD: multiple instructions, multiple data.

Within the SISD class are traditional uniprocessor computers, a single instruction stream operating on a single stream of data. SIMD machines have a single instruction stream, but multiple data streams. This class ranges from single processor vector machines, where a single (vector) instruction initiates action on multiple pieces of data, to machines having multiple identical processors, which all execute the same instruction stream but on different data streams. MIMD machines include the multiprocessor shared-memory vector supercomputers, where each processor has its own instruction stream and works on different streams of data.

Unfortunately, Flynn's classification, although commonly used, is quite restrictive when discussing parallel-architecture computers. There have been several attempts to formulate more detailed classification schemes for the great variety of parallel computers now available. None of these efforts have been entirely successful, and none appear to be in general use. A discussion of representative machines from some of the more common classes follows.

Coarse-Grained Parallelism. The coarsest grained parallelism is the situation in which separate networked computers cooperate to work on a single problem. The computers themselves could be parallel-architecture machines, but that is not important. Communication between the computers is by a network. All computers function as fully autonomous units, perhaps even running different operating systems. As data transmission through a network is usually much slower than transmission through a memory bus, this model will only be effective when relatively large chunks of the problem can be given to individual computers. If the problem requires a great deal of interprocessor communication, the computational throughput could degrade to a level limited by the speed of the network (see NETWORKS).

Many institutions have hundreds, or even thousands, of powerful work stations that are idle for much of the day. There is often vastly more power available in these machines than in any supercomputer center, the only problem being how to harness the power already available. There are network load-distribution tools that allocate individual jobs to unused computers on a network, but this is different from having many computers simultaneously cooperating on the solution of a single problem.

Currently the most widely used commercially available product to support this paradigm is LINDA (20), which maintains a distributed name space for the variables in a network program. Although LINDA successfully hides much of the complexity of network computing from the user, performance is critically dependent on the nature of the problem. Problems that can be decomposed into large, semi-independent portions will generally do well, whereas problems in which the inherent parallelism is finer grained are less likely to exhibit good performance.

Another notable implementation of multiple independent computers is the LCAP (loosely coupled array of processors) project of IBM (21). This system ties together four multiprocessor IBM 3090 Vector Facility computers in a ring topology. In addition to the local memory associated with each processor, a bank of dual-ported memory is shared between neighboring processors. This system exhibits very good performance characteristics on problems that could be mapped onto a coarse-grained architecture, but offers lesser performance when the intermachine communication requirements become greater. Interprocessor communication within a given computer can be done by fast shared memory, but communication time between processors in different computers is comparatively lengthy. Optimization for such a system involves dividing the problem into coarse-grained portions to execute on the different computers, further subdividing those portions into portions that will execute on different processors within the same computer, and then optimizing those individual portions for vector performance. Given the rather extreme costs of such a system, and the challenges of optimization, LCAP has been a valuable research tool rather than a commercial success.

The observation that certain kinds of parallel-computing architectures best support only certain kinds of problems seems to be general. The further observation that interprocessor communication can be the primary impediment to parallel performance is also general. As of this writing, any hope of a truly general purpose parallel computer seems to be remote. The best hope may lie in software efforts that describe problems at higher levels of abstraction, which can then be ported and optimized for different parallel architectures (22).

MIMD Multicomputers. Probably the most widely available parallel computers are the shared-memory multiprocessor MIMD machines. Examples include the multiprocessor vector supercomputers, IBM mainframes, VAX minicomputers, Convex and Alliant minisupercomputers, and Silicon

Graphics server machines. Most of these computers have several CPUs sharing a common memory and usually a common bus. The shared memory, common bus architecture is both the best and most limiting feature of these machines. It is often quite easy to implement parallel programs on these machines, sometimes as simple as using a compiler option directing the compiler to look for opportunities for parallelism within the code. It is unlikely, though, that peak performances will be achieved with so little human intervention and, in general, programs must be restructured to expose the parallelism inherent in the problem to the compiler. For those machines with vector instructions, multilevel optimization must be performed both for vector performance and for parallel performance. Because all processors share a common memory subsystem, and perhaps also a common path to that memory, the relative efficiency of these computers declines as more processors are added. In extreme cases, throughput may actually decrease as processors are added.

Such a decrease in efficiency, but nevertheless increasing throughput, is illustrated by the following data from LINPACK for the Cray Y-MP 832 (3).

Number of processors	MFLOPS, observed	MFLOPS, peak	Efficiency, %
1	324	333	97.3
2	604	667	90.5
4	1159	1333	86.9
8	2144	2667	80.3

This decreasing efficiency is a general characteristic of shared memory, shared bus computers. This example shows unusually high efficiency compared with many other programs. This may be because LINPACK is such a common benchmark that much effort has been devoted to optimizing it for both vector and parallel computers.

Crossbar Systems. Several systems have been built using crossbar architectures for connecting memory and individual processors. Although a crossbar does not maintain dedicated electrical pathways between all points, it can establish such paths as needed. The BBN Butterfly (23) and later Monarch (24) exemplify this approach. The Monarch consists of as many as 65,536 (²¹) processors communicating with half that number of memory modules linked by a pipelined crossbar switch. Although each processor has six independent functional units, the instruction word format only allows for initiation of two operations per cycle. There is no local memory associated with each processor; all memory is shared memory accessible through the crossbar. In order to avoid the complexities associated with cache coherency, the Monarch processors do not have a local data cache, although they do have a local instruction cache. To compensate for the lack of local data cache and memory, the Monarch processors each have 64 registers. The delays associated with accessing global memory through the crossbar are significant, more than 1 μs; however, the effects of this latency can be minimized by performing other instructions, for which operands are already available in the registers, while awaiting completion of a memory access. The memory subsystem is designed to spread data among as many individual memory modules as possible. An interesting feature of the machine is the ability for every processor to simultaneously read from a single memory location in one cycle. Mutual exclusion is implemented by giving individual processors the ability to gain exclusive access to a given memory location. The future of the Monarch remains in doubt; perhaps it will become another victim of the killer micros.

Linear Topology. An interconnect strategy for which the hardware cost scales linearly is to conceptually arrange the processors in a line, with a dedicated interconnect between adjacent processors. The full bandwidth of each connection is then dedicated solely to communication between the adjacent cells. The Warp computer (25) implements such an architecture, with the two ends connected via an interface unit to a host computer. The original system combined 10 identical 10-MFLOPS processors for a combined peak performance of 100 MFLOPS. Each cell could transfer 80 megabytes per second to and from its neighboring cells, in addition to 20 million 16-bit addresses. Clearly the aggregate bandwidth available within the system is much larger than could be obtained from a single shared bus of comparable cost; however, such a system will perform poorly if a given problem is dominated by data transfers between nonadjacent processors.

The original Warp computer exhibited a wide range of performance, depending on the nature of the problem. With ten 10-MFLOPS processors, 100-MFLOPS peak, a 100 × 100 matrix multiply was performed at an observed 79 MFLOPS; whereas averaging a 512 × 512 image to produce a 256 × 256 image ran at 3 MFLOPS, perhaps slower than if the problem had been run on just a single processor. The critical importance of understanding the problem, the machine architecture, and the possible synergies, or lack thereof, between the two should be apparent.

Transputers. At higher levels of connectedness there is a wide variety of parallel computers. A great many parallel computers have been built using INMOS Transputer chips. Individual Transputer chips run at 2 MFLOPS or greater. Transputer chips have four communication channels, so the chips can readily be interconnected into a two-dimensional mesh network or into any other interconnection scheme where individual nodes are four-connected. Most Transputer systems have been built as additions to existing host computers and are SIMD type. Each Transputer has a relatively small local memory as well as access to the host's memory through the interconnection network. Not surprisingly, problems that best utilize local memory tend to achieve better performance than those that make more frequent accesses to host memory. Systems that access fast local memory and slower shared memory are often referred to as NUMA, nonuniform memory access, architecture.

Another characteristic of Transputer systems is that there may be a start-up phase during which the data pass from the host and are distributed throughout the processor array. Only the edge nodes are physically attached to the host system, so the time needed for the data to propagate to all processors is potentially significant. Depending on the nature of the problem, there may also be an additional, time-consuming, final collection phase during which the resulting data pass from the array back to the host.

Because of their relatively low cost and simplicity, Transputer systems can be built readily. Many systems are marketed as application accelerator boards for personal computers and work stations. A single board that turns a standard work station into a "supercomputer" (for one application at least) can be very attractive, especially in application-specific situations (26). Transputer-based systems typically cost between \$1000 and \$200,000, depending on the number of nodes, among other things.

Hypercube and Massively Parallel Systems. The NCUBE computer is one of a large class of hypercube topology computers. A hypercube of dimension *n* contains 2^{*n*} nodes (Fig. 2). Hypercube topology offers significant advantages in parallel-computer design (27). Hypercubes represent a reasonable trade-off between the number of connectors at each node, with the maximum number of interprocessor "hops" a message will need in order to pass from any one node to another being log₂(*n*). As this value grows slowly with *n*, such systems are scalable. Many other kinds of processor topology can be mapped onto a hypercube, with less-connected schemes, a mesh for example, achieved by simply not using existing connections. In order to double the number of processors in a hypercube array, it is only necessary to add one extra connection to each processor. This is a favorable implication for large-scale implementations.

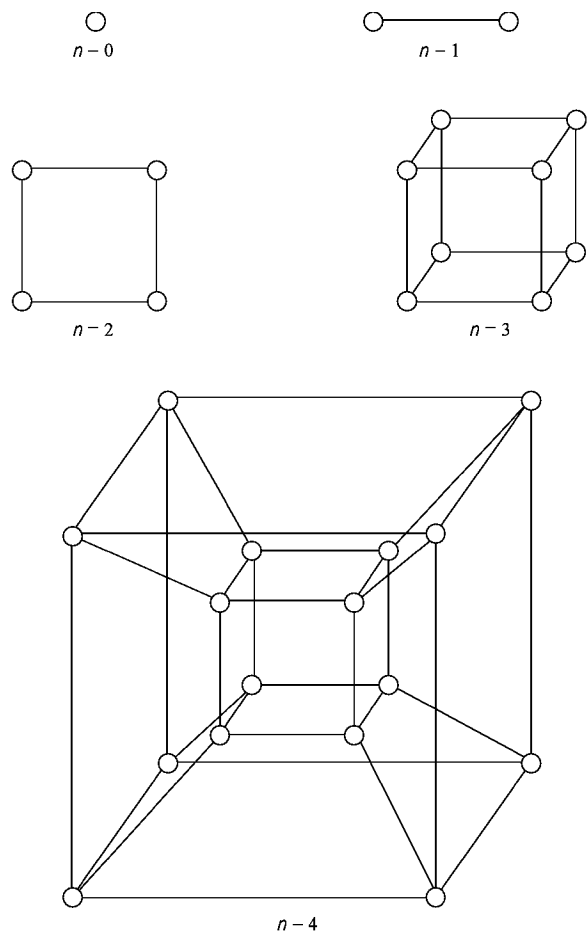


Fig. 2. Hypercubes for $n = 0, 1, 2, 3, 4$ (23).

Courtesy of IEEE.

The NCUBE 2 is a MIMD machine, having between 8 and 8192 processors, each with as many as 64 megabytes of local memory: in an 8192-processor machine there is a maximum of 4 megabytes of memory per processor. All memory is local. A processor needing the value of a variable not already in its own local memory must send a message to the processor whose memory contains that variable. Message passing is performed by dedicated routing units, which means that most of the overhead of message passing is not carried by the computing elements, thereby freeing them to work on computing rather than on message passing. Considerable instruction overlap is possible, so that the latencies involved with remote memory accesses can be overlapped with computation. With each processor rated at 2.4 MFLOPS, an 8196-node system has available a peak speed of 27 GFLOPS in double precision. NCUBE has been a commercial success.

The NCUBE and similar computers often exhibit very good performance on grid-based problems. There is a natural mapping from the spatial nature of the grid to the interconnectivity of the hypercube. Most such grid-based methods do not impose requirements for distant communications; the information needed to update a node value is usually a function of only nearest-neighbor grid points. Finite-element problems that do not perform optimally on vector-architecture machines may show to great advantage on a hypercube MIMD or SIMD system. An example given by NCUBE cites a fluid dynamics study in which a 128-processor system performed 1.8 times faster than a Cray X MP, with every expectation that the problem would scale linearly when implemented on larger hypercubes. Often the parallelism inherent in a problem can be naturally expressed in a way that maps well onto a parallel computer, whereas exposing such parallelism for a vector computer may be quite challenging.

Molecular mechanics simulations can be readily mapped onto such kinds of machine architecture by using the spatial locality of the atoms to determine their allocation to processors. Short-range van der Waals forces can usually be accurately modeled with a cut-off distance of less than one nm, so interprocessor communication requirements can also be localized.

Many problems obviously cannot be decomposed into semi-independent pieces of similar running time. Other problems can be decomposed into subproblems, the running times of which may be quite different and unknown until execution actually occurs. If these program portions are statically allocated to individual processors, or groupings of processors, there exists the possibility for having large parts of the machine idle while waiting for a hot spot to catch up. A great deal of theoretical work is devoted to the problem of dynamic load balancing in MIMD-architecture machines (28).

Whereas the NCUBE and Intel iPSC MIMD computers utilize relatively complex chips at each node, the Connection Machine (29) combines as many as 65,536 simple processors, each of which has 8K or more of local memory (a system maximum of 8 gigabytes in a fully configured system). When operating at full capacity, a Connection Machine can perform 3500 million 32-bit integer operations per second. If equipped with floating-point processors, its peak speed is 21 GFLOPS (double precision). Like the NCUBE, the Connection Machine can be logically subdivided into independent subcubes, thereby facilitating sharing by multiple users and experiments with differing numbers of processors.

The massively parallel approach adopted in the Connection Machine has been termed data parallel. Whereas a uniprocessor must sequentially step through large amounts of data, a data parallel machine moves processors to the data. Aggregate memory to processor bandwidth in the Connection Machine is more than 700 megabytes per second.

The Connection Machine computers have been quite successful, finding application in a wide variety of fields despite relatively high costs. Many image-processing algorithms exhibit good data locality and perform well on a Connection Machine as do grid-based fluid dynamics calculations. With an ability to simultaneously examine independent pieces of data, Connection Machines have found application in full text searching document retrieval areas. Partial differential equations can also be solved with very high efficiency. Protein sequence comparison is another naturally parallel problem well-suited to the Connection Machine. Molecular dynamics calculations have also been adapted to the Connection Machine with good results obtained in the evaluation of nonbonded terms (30).

As of this writing, massively parallel computers such as the Intel Delta, the Thinking Machines CM-2, and NCUBE 2 are yielding as much as 11 GFLOPS on a large-scale variant on the LINPACK benchmark (3). Interestingly, these levels of performance are achieved with much larger problem sizes than used in the traditional LINPACK benchmark.

Application-Specific Hardware. Another interesting trend is the development of application-specific hardware. These systems are usually minimally programmable, but can offer exceptional performance on the class of problem for which they are designed. A particularly interesting application-specific processor for molecular mechanics is currently being developed (31).

Trends. The parallel-computing marketplace is one of rapid change. New companies are developing computers based on new ideas, some

existing vendors are working to enhance the performance of their existing machines, and others are giving up. An end user contemplating a purchase from this market is faced with a bewildering array of choices. Unless the user has only one application, there may be no one computer that is best for all the problems of interest. Additionally, most problems will require at least some effort, and potentially a great deal of effort, to achieve good performance on a parallel computer. Making an *a priori* prediction of the likely performance a given problem might exhibit on a particular parallel machine is often difficult. Although some problems will be obviously analogous to other problems for which good parallel algorithms and experience already exist, at other times there will be no such lead. As vectorization is a kind of parallelism, it is possible that vector-optimized code may also be well optimized for a more parallel computer. Algorithms that had been discarded as inefficient on serial computers are now being found to be optimal for various parallel architectures.

As of this writing, the parallel-processing industry is at an early stage. Further changes and maturing over the years can be expected, including more processors, faster processors, higher bandwidths, larger memories, faster I/O subsystems, and better visualization systems. The most important, but most difficult, advances that must come are in software. In all the cases previously discussed, the ready availability, ease of use, and lowered cost of a given computer, technology has led to a dramatic increase in the utilization of computations running on those kinds of machines. A similar increase should be expected with parallel computers; however, the great breakthrough has not yet arrived. Parallel computing currently remains difficult, often expensive, but frequently very rewarding.

BIBLIOGRAPHY

"Computers" in *ECT* 3rd ed., Vol. 6, pp. 701–719, by T. J. Harrison, International Business Machines Corp.

1. R. W. Hockney and C. R. Jesshope, *Parallel Computers*, Adam Hilger, Bristol, UK, 1981; R. W. Hockney and C. R. Jesshope, *Parallel Computers 2*, Adam Hilger, Bristol, UK, 1988.
2. G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed., Johns Hopkins University Press, Baltimore, Md., 1989, p. 4.
3. J. J. Dongarra, *Performance of Various Computers Using Standard Linear Equations Software*, Oak Ridge National Laboratory, Oak Ridge, Tenn., Aug. 1991.
4. G. Amdahl, *AFIPS Conf. Proc.* **30**, 483–485 (1967).
5. W. J. Here, L. Radom, P. R. Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory*, John Wiley & Sons, Inc., New York, 1986.
6. M. J. Frisch, M. Head-Gordon, and J. A. Pople, *Chem. Phys.* **141**, 189 (1990).
7. S. Obara and A. Saika, *J. Chem. Phys.* **86**, 3963 (1986).
8. G. A. Segal, *Semiempirical Methods of Electronic Structure Calculation, Modern Theoretical Chemistry*, Vols. 7 and 8, Plenum Publishing, New York, 1977.
9. *Journal of Computer-Aided Molecular Design* **4**(1), (1990). Special issue devoted to MOPAC.
10. A. Dinning, *IEEE Computer* **22**, 66 (1989).
11. Attributed to Eugene Brooks.
12. J. J. O'Donnell, *Research and Development*, Trace 300 Series: Technical Summary, Multiflow Corp., Branford, Conn., Jan. 1990, p. 80.
13. D. W. Wall, *Proceedings of the Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*, ACM, 1991, p. 176.
14. B. Rau and co-workers, *IEEE Computer* (Jan. 1989). Special issue devoted to design of high performance computers.
15. J. E. Smith, *IEEE Computer* **22**, 21 (1989).
16. B. Smith, *17th Annual Symposium on Computer Architecture*, Seattle, Wash., May 1990.
17. *Comp. J.* **34**(4), 289 (1991). Special issue on concurrent programming.
18. *IEEE Computer* (June 1990); M. D. Hill and co-workers, *18th Ann. Int. Symp. Comp. Archit.* **19**(3), 298 (1991).
19. M. J. Flynn, *IEEE Trans. Comput.* **C-21**, 948 (1972).
20. N. Carriero and D. Gelernter, *ACM Trans. Computer Systems* **4**(2) (May 1986).
21. D. Folsom, *MOTECC, Modern Techniques in Computational Chemistry*, Escom, Leiden, The Netherlands, 1990, p. 1091.
22. J. C. Browne and co-workers, *Proc. 1989 Int. Conf. Parallel Processing* **2**, 98 (1989).
23. *Butterfly Product Overview*, BBN Advanced Computers, Cambridge, Mass., 1987.
24. R. D. Rettberg and co-workers, *IEEE Computer*, 18 (Apr. 1990).
25. M. Lam and co-workers, *IEEE Trans. Computers* **C-36**, 1523 (Dec. 1987).
26. R. S. Cok, J. Gerstenberger, C. Lawrence, H. Neamtu, and D. S. Cok, *Transputing '91*, Vol. 1, Sunnyvale, Calif., Apr. 22–26, 1991, p. 15.
27. J. P. Hayes and co-workers, *IEEE Micro*, 6 (Oct. 1986); G. C. Fox and P. C. Messina, *Sci. Am.*, 67 (Oct. 1987).
28. H. J. Siegel and M. Jeng, *Proc. 1989 Int. Conf. on Parallel Processing* **2**, 57 (1989).
29. W. D. Hillis, *Sci. Am.*, 108 (June 1987).
30. B. Murray, P. Bash, and M. Karplus, *Molecular Dynamics on the Connection Machine*, Thinking Machines Corp., Cambridge, Mass., 1989.
31. B. R. Brooks and co-workers, Molecular Graphics and Simulation Laboratory, NIH, Bethesda, Md., private communication, Aug. 1991.

General References

IEEE Computer (Sept. 1991). Issue devoted to trends expected into the year 2000.

T. Leighton, *Introduction to Parallel Algorithms and Architectures: Arrays, Trees, Hypercubes*, Morgan-Kaufmann, San Mateo, Calif., 1991.

R. W. Counts, *J. Comp.-Aided Molec. Des.* **5**, 167 (1991).

Ian Watson
Consultant

CONCRETE.

See CEMENT.

CONDIMENTS.

See FLAVORS AND SPICES.

CONGO COPAL.

See RESINS, NATURAL.

CONSTANTAN.

See NICKEL AND NICKEL ALLOYS.

CONSTRUCTION MATERIALS.

See BUILDING MATERIALS.

CONTACT LENSES

Contact lenses are small, hemispherical-shaped optical devices placed in contact with the transparent tissue at the front of the eye called the cornea. The capillary attraction of the liquid tear layer between the contact lens and the cornea and the partial coverage of the lens by the eyelids prevent contact lenses from being dislodged during eye movement and blinking. Contact lenses are used to correct vision deficiencies such as myopia (nearsightedness), hyperopia (farsightedness), presbyopia (loss of near focusing power with age), and astigmatism. An increasingly common use of contact lenses is to change the color of a normal eye or to improve the appearance of an eye disfigured by injury or disease. Contact lenses can also be used as therapeutic devices in the medical treatment of certain eye diseases or injuries.

The concept of a contact lens device for modifying the optical power of the eye was described by Leonardo da Vinci and later by Rene Descartes and Thomas Young. In 1823, Sir John Herschel described the application of a contact lens device specifically for the purpose of correcting vision. The first contact lens was fitted to a human eye for correction of vision in 1888. The early lenses were made of blown or molded glass and were difficult to wear. The contact lens industry began to grow after World War II when the plastic poly(methyl methacrylate) [9011-14-7] (PMMA) was found to be a suitable material for contact lenses in terms of optical properties, biocompatibility, and manufacturability. In the late 1950s a hydrogel material based on poly(2-hydroxyethyl methacrylate) [25249-16-5], poly(HEMA) was discovered. Contact lenses made with this water-containing soft plastic promised significantly improved comfort compared with the hard PMMA lenses. Poly(HEMA) lenses were used in Czechoslovakia in the 1960s; popularization of soft contact lenses in the United States occurred in the 1970s, following the 1971 introduction by Bausch & Lomb of the Soflens contact lens.

The number of contact lens wearers has grown to an estimated 24 million in the United States and 50 million worldwide. Concurrently, there has been a proliferation of contact lens manufacturers and products. The 1980s saw the widespread introduction of lens products made of more oxygen-permeable materials, ie, rigid gas-permeable (RGP) materials that made PMMA lenses virtually obsolete and high water content hydrogels that competed with HEMA-based lenses.

Research continues for materials having improved oxygen permeability, deposit resistance, and comfort. In addition, there is the search for breakthrough lens designs, such as bifocal contact lens products. The challenge for contact lens manufacturers is to capture more than a 15% share of the vision correction population, now dominated by traditional eyeglasses. Furthermore, competition includes developing new surgical techniques, eg, excimer laser refractive surgery that modifies the shape of the cornea to correct vision, potentially without the use of eyeglasses or contact lenses. Detailed information on historical and clinical aspects of contact lenses (1,2) and overviews of refractive surgery techniques (3,4) are available.

Clinical Aspects.

No particular contact lens type or product is considered universally superior. In some regions of the world hard lenses dominate the market, eg, some European countries and Japan; in other regions, eg, North America and Scandinavia, soft lenses dominate. Contact lens practitioners select their preferred type of lens using criteria other than just lens material properties. However, among soft lenses, HEMA-based lenses are prescribed most often, and among hard lenses, silicone–acrylate RGP lenses are most common.

To remain safe and efficacious on the eye, contact lenses must maintain clear and wetted surfaces, provide an adequate supply of atmospheric oxygen to and adequate expulsion of carbon dioxide from the cornea, allow adequate flow of the eye's tear fluid, and avoid excessive abrasion of the ocular surface or eyelids, all under a variety of environmental conditions. The clinical performance of a contact lens is controlled by the nature of the lens material; the lens design; the method and quality of manufacture; the lens parameters or specifications prescribed by the practitioner; and the cleaning, disinfection, and wearing procedures used by the patient.

Corneal edema occurs with almost every type of contact lens as a result of corneal hypoxia during lens wear. It is most noticeable when lenses are worn during sleep. Although corneal edema is reversible, excessive and prolonged edema can lead to changes in corneal curvature, growth of blood vessels into the central cornea across the line of sight, or vision reduction through corneal haziness. Excessive edema may predispose the cornea to abrasion,

inflammation, or infection, leading to a corneal ulcer. Consequently, a primary goal of contact lens manufacturers has been to develop materials with higher oxygen permeability to reduce corneal hypoxia. For a lens 0.1 mm thick, an oxygen permeability coefficient (Dk) of approximately 100 barrer units is thought to be sufficient to produce no more corneal edema than occurs when the eyes are closed during sleep with no contact lens in place (5,6).

All contact lenses tend to accumulate debris, deposits, and discolorations on and sometimes below the surfaces (7). The deposits are usually proteinaceous or lipid but may also be inorganic compounds and microorganisms, eg, fungi and bacteria. The source of these deposits is usually the eye itself, the tear film that bathes the surface tissues, the eyelid glands, the mucoid substance secreted by certain conjunctival cells, or the immunoglobulins secreted through the vascular system. Although regular cleaning and disinfection can control these accumulations and reduce the likelihood of clinical problems such as reduced vision, inflammation or red eye, abrasion, and infection, the deposits are considered undesirable (8,9).

Clinical experience has shown that certain types of lens materials are more prone to deposit problems. In general, lenses with negatively charged moieties at the surface accumulate greater amounts of lysozyme, the principal tear film protein (10). The introduction and use of disposable lenses make these deposits and their clinical problems less significant.

Classification of Contact Lenses. All contact lenses can be divided according to the wearing modality. Daily wear lenses are worn during the day only, being placed on the eye in the morning and removed, cleaned, and disinfected before sleep at night. Extended wear or flexible wear lenses can be worn continuously for several days and nights, including during sleep. They have a higher oxygen transmissibility than daily wear lenses, achieved either by making the lens thinner or by making the lens from a more oxygen-permeable material.

Contact lenses can be divided further according to the type of vision deficiency for which they provide optical correction, ie, myopia, hyperopia, presbyopia, and astigmatism. A toric lens is designed to correct astigmatism; an aphakic lens is designed to correct the high amount of hyperopia created when the natural crystalline lens of the human eye is surgically removed because of a cataract. A bifocal lens or multifocal lens is designed to have at least two distinct optical segments to correct vision, usually one for distant and one for close vision. For some bifocal lenses and for all toric lenses, the lens is kept on the eye in one orientation, without rotating off axis, by designing the lens with some ballasting. This provides the optimal correction of vision. Because their orientation on the eye is not critical to optical performance, nonbifocal and nontoric lenses have rotationally symmetric optics.

Contact lenses can be subdivided further according to whether they are tinted and or disposable and by the material from which they are formulated. In addition, contact lenses can be used medically for the treatment of certain eye diseases and injuries by providing a protective cover while the tissue heals itself or by acting as a drug delivery system; such lenses are termed therapeutic lenses, bandage lenses, or shields.

Disposable lenses and lenses used in planned replacement programs are the most recent contact lens development. Traditionally, the usable life of contact lenses has been one year or longer because of the use of various lens-cleaning procedures. Disposable lenses were introduced to avoid the inconvenience and likelihood of patient noncompliance in carrying out such cleanings. They were originally intended to be used once, for one week of continuous wear, discarded, and replaced with a new pair. Disposable lenses are now prescribed with a variety of different wearing schedules, disposal periods, and cleaning and disinfection procedures. Disposable lenses are not intrinsically different in design or material from nondisposable lenses, but are generally packaged in foil-blister packets whereas nondisposable lenses are packaged in glass vials. Although the single lens cost is lower for a disposable lens, the total cost to a patient of a disposable lens system, eg, 52 lens pairs per year, is somewhat higher than the cost for a nondisposable lens system, eg, 1–2 lens pairs per year, including solutions for cleaning and disinfecting the lenses. By necessity, disposable lenses are manufactured using cost-efficient manufacturing technology to produce the huge quantities needed for the marketplace.

It is convenient to classify lenses according to whether they are rigid (hard) or flexible (soft). Some newer materials give rise to lenses that can be termed semirigid, semisoft, and the combination soft–rigid. Hard lenses, including RGP lenses, are invariably nonhydrogels, although their surfaces are wettable on the eye. Soft lenses are usually hydrogels. One notable exception is the silicone elastomer lens that is quite flexible but is not a hydrogel.

The flexibility of a contact lens material dictates, to a large extent, the design of the lens. Hard lenses are generally 8–10 mm in diameter, cover only some of the corneal surface, and fit mostly in between the upper and lower eyelids. Soft lenses are generally 13–15 mm in diameter, cover the entire corneal surface, extend onto the white of the eye, and fit well underneath the upper and lower eyelids. Hard lenses have slightly superior optical performance on the eye, superior oxygen permeability (for some rigid materials), enhanced flow of the liquid tears and removal of ocular debris from between the lens and the cornea, and durability. In addition, hard lenses allow easier cleaning and disinfection methods for the patient in some situations. Soft lenses have considerably greater comfort when first placed on the eye and more secure positioning under the eyelids, which prevents accidental dislodgement.

Properties of Contact Lenses

Not every polymer can be manufactured successfully into a contact lens. Several important properties for both ocular physiology and patient handling are required of a material for a contact lens application (1,11). In addition, the type of lens application, ie, rigid, flexible, or soft, will dictate the range and importance of the key properties.

Oxygen Permeability. The most critical factor for a contact lens is oxygen permeability. With no blood vessels, the cornea must obtain oxygen from the atmosphere; the cornea will swell and become hypoxic with poor oxygenation. A contact lens has the potential to act as an oxygen barrier to the cornea, and although the cornea is oxygenated partially from a tear pump mechanism by which air-oxygenated tears flow under the contact lens, this amount of oxygen is not sufficient to allow for wearing times longer than a day. Oxygen permeability (Dk) is an intrinsic property of the material, and oxygen transmissibility (Dk/L , where L is the lens thickness) depends on the thickness of the material. There is also a growing concern for the measurement of carbon dioxide permeability; buildup of carbon dioxide behind the lens may cause significant pH changes in the tear (12).

In the contact lens field, barrer units are typically used for the oxygen permeability coefficient (Dk).

1 barrer

$$\frac{10^{-11} \text{ cm}^3 \text{O}_2 \text{ (at STP) } \cdot \text{cm}}{\text{cm}^2 \cdot \text{s} \cdot \text{mm Hg}} \qquad \frac{0.335 \text{ mmol}}{\text{m} \cdot \text{s} \cdot \text{TPa}}$$

See BARRIER POLYMERS for a discussion of permeability units.

Oxygen permeability can be measured by a variety of methods. Classical techniques such as ASTM D1434 measures the amount of oxygen transported across a polymer membrane when an infinite gradient is established (13). An alternative-method measures the oxygen volumetrically (14). The most commonly used technique is a polarographic method (15) adapted for contact lenses (16). Although several variations on the method are used, the basic principle is consistent. The lens polymer is placed over a gold electrode and immersed in a saline solution. When a potential is placed across the electrode and the solution, oxygen arriving at the gold electrode is reduced to hydroxyl and the subsequent current is measured with a sensitive ammeter. When the current at the gold electrode reaches a steady state, the current is directly related to the oxygen flux through the polymer. Oxygen permeability can be calculated as the slope of the line of the measured flux plotted against the sample thickness. Although a value can be determined from a single measurement on a calibrated system, measurement of oxygen flux for multiple thicknesses provides a more accurate value of Dk (17). The measurement of oxygen permeability has been frequently studied over the last several years and a number of modifications to the calculations have been made. The most significant of these is the correction for edge effects (18,19). Because the oxygen flux is not one-dimensional through the material, some account must be made for the oxygen coming from the edges of the sample.

The physiological effect of a particular lens can be determined by measuring the increased thickness of the cornea after lens wear; studies have shown the relationship between the Dk/L of a lens and the subsequent swelling of the underlying cornea (20,21).

Oxygen permeabilities for currently marketed hydrogel materials can be directly related to water content. Figure 1 shows the manufacturer-reported values of Dk and water content for marketed hydrogel materials (22). For rigid and flexible materials the Dk is an indirect function of siloxane content. Polydimethylsiloxane has one of the highest measured oxygen permeabilities of all tested materials. Incorporating siloxanes into a hydrophilic matrix is an important research area of contact lens manufacturers, which may lead to the development of novel experimental hydrogels that exhibit low water content but high Dk values.

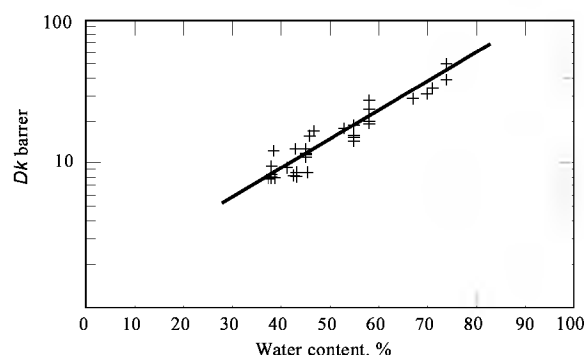


Fig. 1. Correlation of Dk with percent water (log-linear plot). 1 barrer = 10^{-11} cm³ O₂ (at STP)·cm (cm²·s·mm Hg).

Water Content and Refractive Index. The water content of a hydrophilic contact lens is a determinant of other properties. The relationship of water content and Dk is discussed above. Water content in lenses is inversely related to refractive index (23), a key property for vision correction. A lens material with a higher refractive index refracts light to a greater degree, allowing more vision correction with a thinner material. The water content of a lens is generally determined gravimetrically or inferred from the relationship to refractive index, measured with a refractometer (24).

Water content indirectly affects other lens characteristics. Water evaporation from the lens can result in a dry eye sensation and subsequent desiccative erosion of the cornea. Clinical studies have shown the incidence of corneal erosion as a result of lens desiccation to be a material-dependent and water-content-dependent phenomenon (25,26). The nature of water and sodium ions in hydrogels has been studied primarily by nmr and thermal techniques (27,28). An empirical relationship between water mobility in contact lens polymers and desiccative staining has been proposed (29).

Mechanical Properties. For hydrogel materials, the most important mechanical properties are strength and modulus. The modulus of a material determines relative stiffness and is generally measured under tensile conditions. A low material modulus, 0.4 MPa (<40 g mm²), handles more poorly outside the eye and possesses poor drapability; however, the material is generally more comfortable and fits well. A high modulus lens, between 1 MPa and an upper limit of approximately 2 MPa (100–200 g mm²), handles much better and stands up on a fingertip, but it may irritate the conjunctiva of the eye or interact with the lid to cause discomfort. Generally, a higher modulus lens requires more parameters to fit an entire population of wearers. Modulus typically is measured using ASTM D1708 (30) modified to accommodate the smaller samples with a variety of configurations; the sample should remain in the hydrated state for an accurate value. The test can be carried through until the sample breaks and the tensile strength and elongation of the material are obtained. Naturally, the modulus is an intrinsic property and the actual lens properties depend on thickness.

The strength of a hydrophilic lens material determines how well the material survives the patient's handling, cleaning, and disinfection regimen. Materials with higher Dk have been found to be generally more fragile (31).

Although the strength of a lens material can be inferred from the tensile strength, the property usually measured is tear strength. Typically determined by ASTM D1938 (32), or modifications thereof, a small tear is initiated and the force required to continue the tear is measured. Tear strengths for soft contact lens applications are quite low; some marketed lenses have tear strengths under 19.6 N m (2 g mm). Flexible lens materials are typically elastomers and have higher moduli and significantly higher tear strengths than hydrogel materials, determined by using similar methodology.

Rigid lenses have tensile properties in the tens of MPa (thousands of g mm²), which give rise to excellent vision but poorer comfort. Although tensile properties can be determined for these materials, lens performance is not as sensitive to these values as for hydrophilic lens materials.

Flexure testing has also been used to measure the strength of a RGP material (33). The strengths measured from flexure tests can be empirically related to a material's resistance to breakage during typical lens handling.

Other mechanical properties determined for RGP materials dictate their ease of processing. Glass-transition temperature (T_g), or softening temperature, also will dictate processing conditions, ie, curing, annealing, and lathing, but must be greater than eye temperature for the lens to remain rigid and maintain optics. Material hardness, measured on either the Shore or Rockwell scale, determines the capability of a material's lathing.

Light Transmittance. Contact lenses should transmit light in the visible range from 400–800 nm. The addition of tints affects light transmittance, but a majority of tinted lenses still transmit more than 85% of visible light (34). Of growing interest is the absorption of light in the ultraviolet region; there is concern about the possible relationship of the transmittance of uv light to cataracts and ocular health (35,36). Companies have developed uv-blocking lenses that contain an additional lens component that absorbs uv light between 300 and 400 nm (37,38). No absorber will remove 100% of uv light because leaking of absorbance into the visible region gives the lens an undesirable yellow tint. Absorbances are measured with standard spectrophotometers; techniques differ in aperture of the light beam and placement of the sample in the beam.

Wettability. Wettability dictates the biocompatibility of a contact lens material in the ocular environment. A nonwetting material will not maintain a uniform tear film between blinks. The result is a lens that is uncomfortable, provides poor vision, and has the potential for deposit formation. The most common method for wettability measurement is contact angle. Two standard methods for determining this value exist: the sessile drop technique, ie, a drop of solution is placed on the sample and the angle formed by the material surface and the leading edge of the drop is measured with a goniometer; and the captive bubble technique, ie, the sample is immersed in a solution, a bubble is captured beneath the surface, and the angle between the material and the leading edge of the air bubble is measured with a goniometer (39). A more recent technique is the measure of dynamic contact angle using the Whilhelmy plate technique (40), ie, the force of driving a sample into solution is measured to obtain the advancing contact angle, equivalent to the sessile drop value, and the force to withdraw a sample from the solution is measured to obtain the receding contact angle, equivalent to the captive bubble value (41). Contact angles have also been measured *in vivo* (42). These methods are used extensively by researchers and contact lens manufacturers, but contact angle values are not always indicative of on-eye wettability. For example, the contact angle of PMMA determined by the sessile drop technique (43) is >60, which indicates a poorly wetting surface. Yet the lens has demonstrated excellent *in vivo* wettability.

Wettability is believed to be a dynamic process involving interactions of the ocular environment and lens material (44). No method has been found to predict accurately the *in vivo* wettability of a lens material.

Deposition. Lens spoilage from tear film deposition is a common reason for lens replacement and mandates the frequent cleaning of nondisposable contact lenses. When deposits become significant, they affect vision, lens properties (45), and ocular health (46). Deposits that form on contact lenses are complex and contain proteins, lipids, and minor components such as calcium (47,48). The mechanism of deposit formation is of research interest and draws on work in the biomedical device field (49,50). Studies have shown that for hydrophilic lenses deposition depends on material; polymers containing a charged species show greater deposition than uncharged polymers (51,52). Although no standardized tests exist for the measurement of deposition, each manufacturer has empirical tests to screen materials for protein or lipid deposition potential. These tests are not 100% predictive, because actual lens deposition is a function of lens type, patient tear chemistry, cleaning and disinfection regimen, and environment.

Although a variety of test methods, eg, Dk , modulus, and tear strength, exist to determine key properties of potential contact lens materials, a number of properties, eg, wettability and deposition, have no predictive methodology short of actual clinical experience.

Rigid (Hard) Lenses

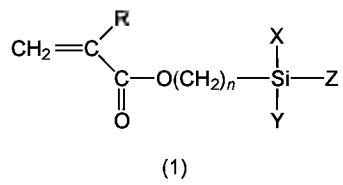
Hard lenses can be defined as plastic lenses that contain no water, have moduli in excess of 5 MPa (500 g mm²), and have T_g well above the temperature of the ocular environment. Poly(methyl methacrylate) (PMMA) has excellent optical and mechanical properties and scratch resistance and was the first and only plastic used as a hard lens material before higher oxygen-permeable materials were developed. PMMA lenses also show excellent wetting in the ocular environment even though they are hydrophobic, eg, the contact angle is 66°.

PMMA hard lenses received wide acceptance when introduced in 1945 and thrived throughout the 1970s. The primary drawback of PMMA lenses is lack of oxygen transmissibility; the little oxygen that the cornea receives during PMMA lens wear is almost exclusively through tear exchange under the lens. During the 1980s, a series of hard lenses with higher oxygen permeability, the RGPs, were developed and introduced to the market. As a result, the popularity of PMMA lenses decreased even further.

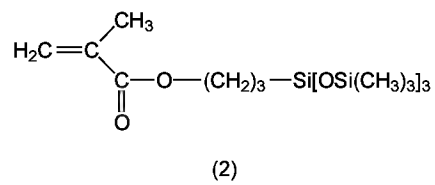
Early RGP Lenses. Because of poor oxygen permeability, PMMA lenses can only be worn for daily wear. To overcome the poor oxygen permeability of PMMA lenses, other rigid polymers have been investigated for contact lens application. One of the first polymer families to be successfully developed was the cellulose acetate butyrate [9004-36-8] (CAB) lenses (53,54), which showed oxygen permeability of 10–20 barriers, approximately two orders of magnitude increase in oxygen permeability than that of PMMA lenses. However, compared with PMMA lenses, CAB lenses had a higher tendency to warp, scratch, chip, and discolor. Another class of polymers evaluated for use as a hard lens material was styrenics. Poly(*p*-*t*-butylstyrene) [26009-55-2] has a higher free volume in the three-dimensional packing, due to the bulkiness of the *t*-butyl group, giving oxygen permeability of 25 barriers; this value is considered to be high among the polymers evaluated for contact lens application in the early 1970s. Poly(*p*-*t*-butylstyrene) was commercialized as the Airlens contact lens by Wesley-Jessen. Although possessing good flexural strength, the material showed poor scratch resistance.

Advances in RGP Materials. Silicone rubber is an attractive contact lens material (55) with a high oxygen permeability, resulting from the flexibility of Si–O–Si bonds. Since the mid-1970s, a tremendous amount of research-and-development work has been performed with rigid gas-permeable lens materials employing the high oxygen permeability of siloxanes. These efforts have led to the introduction of many highly oxygen-permeable hard lenses.

Silicone Acrylates. The development of rigid gas-permeable lens materials advanced significantly after the development of polysiloxanylalkyl acrylates and methacrylates (1), as a component in hard lens materials (56,57), as claimed in a series of patents (58–62).



where $\text{Z} = \text{A} - \left[\begin{array}{c} \text{A} \\ | \\ \text{Si} - \text{O} \\ | \\ \text{A} \end{array} \right]_m -$, X, Y alkyl, phenyl or Z, A alkyl or phenyl, and R H or CH₃; The simplest structure of this class of monomers, 3-(methacryloxy)propyl tris(trimethylsiloxy)silane, [17096-07-0] (TRIS) (2), gives polymer with oxygen permeability >200 barrers and is particularly useful as a component for hard lens application.



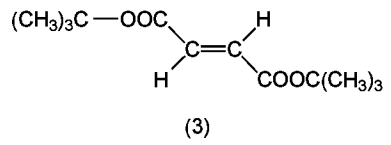
Unlike a typical polysiloxane, which imparts high oxygen permeability from the flexibility of the dimethylsiloxane units, the high oxygen permeability of TRIS-containing polymers may be the result of the bulkiness of tris(trimethylsiloxy)silane units, which gives a high free volume of polymer available for oxygen transport.

The preparation of silicone acrylate–methacrylate monomers has continued with the purpose of obtaining highly oxygen-permeable hard lens materials. Bulky polysiloxanylalkyl diacrylates–dimethacrylates were modified (63,64). Polysiloxanylalkyl acrylates–methacrylates modifications include changing the arrangements of alkyl units (65,66) attached to the siloxane groups and making the siloxane units bulkier (67,68). Silylmethylene methacrylates were formed by replacing siloxane units with silane units (69,70). Acrylamide–methacrylamide analogues of polysiloxanylalkyl acrylate–methacrylate also were prepared (71). Owing to the availability of monomers, the ease of preparation, and the excellent physical properties, acrylates–methacrylates are the favored choice for a hard lens material.

During the 1980s, unsaturated functional groups other than acrylate–methacrylate were investigated as alternative polymerizable groups to provide hard lens materials with high oxygen permeability by incorporating polysiloxanylalkyl groups. These unsaturated functional groups include vinyl (72–74), vinyl carbamate–carbonate (75), styrene (76,77), itaconate (63), alkyne, and fumarate. Among these, alkyne- and fumarate-based systems were investigated most extensively and are described below. Because of the difference in molecular structure and the ability to polymerize, these unsaturated functionalities offer lens materials with unique morphology not achievable with acrylate systems.

Polyalkynes. Poly(1-trimethylsilyl)propyne (PTMSP) was prepared (78) and found to have unique properties, when compared with polyacetylene, ie, mainchain double bonds are not conjugated (79). PTMSP is colorless, soluble in common organic solvents, amorphous, and has the highest gas (including oxygen) permeability of any known polymer. The high gas permeability (oxygen permeability of approximately 10,000 barrers) is likely the consequence of a large excess free volume caused by this polymer’s unique structural features (80). However, the conformational changes that cause tighter packing of silane moieties in the solid state reduce oxygen permeability. As a result, the polymer was not further evaluated as a hard lens material.

Polyfumarates. Because of steric hindrance, fumarates are difficult to polymerize into high molecular-weight polymers. However, it was demonstrated during the early 1980s that bulky fumarates, such as di-*t*-butyl fumarate, C₁₂H₂₀O₄ [7633-38-7] (3), can be polymerized thermally into high molecular-weight polymers (81–84).



These polymers have good optical quality as well as high oxygen permeability; their application in the optical area has been cited (85). Subsequent studies on fumarate polymerizations led to a series of interesting systems useful as hard lens materials. For example, bis(siloxanylalkyl) fumarates, which are TRIS analogues, can be polymerized–copolymerized thermally to give polymers with oxygen permeability as high as 480 barrers (86,87), potentially higher than that of TRIS polymer. Compositions containing one dialkyl fumarate and either one bis(siloxanylalkyl) fumarate or a fluoroalkyl fumarate are also useful as hard lens materials with an oxygen permeability as high as 140 barrers (88). Fluoroalkyl fumarates copolymerized with fluoroalkyl itaconate (89) or fluoroalkyl mesoconate (90), and fumarates copolymerized with methacrylates, also give hard lens materials (91–93).

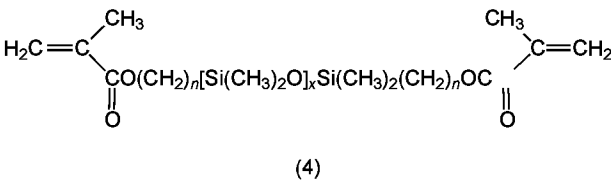
Most of the silicone-based monomers claimed useful for hard lens applications also have been evaluated as components in silicone hydrogels targeted for contact lens applications.

Fluoroalkyl Acrylates–Methacrylates–Itaconates–Mesoconates. Fluoroalkyl acrylates have been copolymerized with alkyl acrylates

to give copolymers with increased oxygen permeability useful as a hard lens material (93). However, the principal role of fluoroalkyl acrylate–methacrylate in hard lens material has been to provide lipid resistance. To obtain a highly oxygen-permeable lens material, monomers with polysiloxane–bulky siloxane fragments are necessary. Polysiloxanes are hydrophobic and thus contact lenses manufactured from silicone monomers naturally absorb lipids from the tear fluids and exhibit serious wettability problems. Fluoropolymers have been known to have low surface energy and resist foreign materials. By incorporating fluoroalkyl acrylates–methacrylates–itaconates–mesoconates, silicone-containing hard lenses possess high oxygen permeability and better lipid resistance (60,62,87,88,94–102).

Other RGP Components. All hard lens materials, regardless of the oxygen permeability, must meet certain requirements, ie, the polysiloxanylalkyl monomer provides oxygen permeability and the fluoroalkyl monomer provides lipid resistance. Unlike PMMA lenses, and because of the hydrophobicity of silicones, hard lenses containing polysiloxane units need a significant amount of hydrophilic monomer to maintain sufficient wettability for patient wear. Although nonionic hydrophilic monomers are mentioned in the hard lens materials patent literature, an ionic wetting monomer, such as acrylic acid [79-10-7] or methacrylic acid [79-41-4] at 1–15%, generally is required for a silicone-based hard lens to maintain sufficient wettability; ionic monomers are cited in most hard lens material patents.

Because of high glass-transition temperatures (T_g), hard lenses are manufactured by thermal polymerization of the monomers, and thermal catalysts are a necessary part of an RGP lens formulation. One dimethacrylate or diacrylate generally is used as a cross-linker in a hard lens composition to help maintain dimensional stability. This kind of cross-linker also may affect other properties, such as mechanical strength. When a polyfluoroether-based diacrylate or dimethacrylate is used, lipid resistance and oxygen permeability improve over the common acrylic monomers (103). When a polysiloxane-based diacrylate or dimethacrylate, such as α , ω -bis(methacryloxyalkyl) polydimethylsiloxane (4) where $n = 1-10$ and $x = 1-200$, is used, lens material with high oxygen permeability and toughness is obtained (104).



Rigid Gas-Permeable Lens Products. From the late 1970s to the early 1990s, a long series of RGP lens products were introduced, offering higher oxygen permeability, better wettability, and more deposit resistance. Table 1 gives a representative list of RGP products introduced to the marketplace. Depending on the method of measurement, a disagreement in published oxygen permeability values exists (6), eg, an oxygen permeability as high as 110 barrers was reached with a fluorosilicone acrylate system. Clinical results suggest that fluorosilicone acrylate hard lenses offer better deposit resistance as well as good oxygen permeability (105).

Table 1. Partial List of Rigid Gas-Permeable Lens Materials

USAN name ^a	Material composition	Dk , barrer ^b	Trade name	Manufacturer
airfocon	<i>t</i> -butylstyrene ^c	25	Aidens	Wesley-Jessen
fluorofocn	fluoropolymer	95	Advent	3M Corp.
itafluorofocn	fluorosilicone	55	Equalens	Polymer Technology
itafocon	silicone acrylate	14	Boston II	Polymer Technology
		26	Boston IV	Polymer Technology
		32	Optacryl K	Optacryl
		54	Optacryl EXT	Optacryl
		67	Optacryl Z	Optacryl
pasifocon	silicone acrylate	49	Paraperm EW	Paragon
		92	Paraperm EW II	Paragon
siflufocon	silicone acrylate	55	Quantum I	Bausch & Lomb
		100	Quantum II	Bausch & Lomb

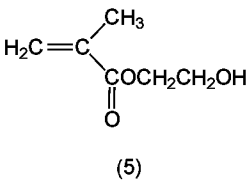
^a U.S. Adopted Name Council designation for polymeric materials.
^b 1 barrer = 10^{-11} cm³ O₂ (at STP)·cm/s·cm²·mm Hg.
^c *t*-C₄H₉–C(H)₄–CH=CH₂ [25338-51-6]; homopolymer [9053-30-9].

The development of rigid gas-permeable lenses resulted in tremendous growth during the 1980s. However, after replacement of most PMMA lenses with rigid gas permeables, growth of this business slowed significantly, attributed to unsatisfactory lens wettability and poor comfort. Wetting conditioning solutions have been investigated to reduce the wettability problem. Ultimately, more wettable lens materials need to be developed. The lens comfort issue may partially be resolved with the development of a variable modulus lens (106), achieved by gradually decreasing the lens modulus from the optical zone to the peripheral region.

Hydrogel Contact Lenses

Hydrogels are water-containing polymers, hydrophilic in nature, yet insoluble. In water, these polymers swell to an equilibrium volume and maintain their shape. The hydrophilicity of hydrogel is a result of the presence of functional groups such as –NH₂, –OH, –COOH, –CONH₂, –CONH–, –SO₃H, etc. The insolubility and stability of hydrogels are caused by the presence of a three-dimensional network. The scope, preparation, and characterization of hydrogels has been reviewed (107).

Hydrogels, ie, gelatin and agar, have been known for a long time. In the late nineteenth century, Herschel proposed the use of jelly materials on the cornea for the correction of vision (108). In 1960, the use of synthetic hydrogels for contact lenses was proposed and several U.S. patents were obtained for the invention of cross-linked hydrophilic polymers, eg, systems based on 2-hydroxyethyl methacrylate [868-77-9] (HEMA) (5) (109–112).



In 1971, Bausch & Lomb received U.S. Food and Drug Administration approval for a soft contact lens based on the HEMA system and launched Soflens. The soft lens market has grown dramatically; soft lens surpassed hard lens wear by the late 1970s, accounting for about 80% of the U.S. contact lens market in the early 1990s.

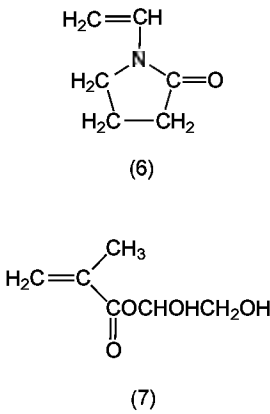
Polymers used in soft contact lenses have not been limited to those derived from HEMA; other hydrophilic monomers also were explored and used. From the late 1960s through early 1970s, the primary concern among investigators was to develop proprietary formulations with higher water content and thus higher oxygen permeability.

Hydrogels used for contact lens applications can be classified based on the chemical structure of hydrophilic monomers and on the water content of the hydrogel lens. These classifications are not unique because the design of hydrogels is based on the performance desired; in most cases, more than one component is used for the same or complementary purpose. In fact, some hydrophilic monomers are used in both high water and low water lenses.

The U.S. Food and Drug Administration, which has the functional authority to regulate contact lenses, classifies soft contact lens according to the water content and ionic character of the soft lens (113), ie, group I, low water (<50%) nonionic; group II, high water (>50%) nonionic; group III, low water (<50%) ionic; and group IV, high water (>50%) ionic. This classification is particularly useful for clinicians and the FDA in determining care system efficacy.

Typical Hydrogel Components. Regardless of water content, a hydrogel lens formulation consists of hydrophilic monomers, cross-linkers, and initiators. Hydrophilic monomers provide the water needed in a hydrogel, and the cross-linkers help to hold the hydrogel's shape. The initiator starts the polymerization and curing and is removed from the hydrogel after the polymerization. In addition to the above three components, a hydrogel formulation also may contain a hydrophobic monomer as a physical property modifier and a solvent to facilitate the processing.

The water content of a hydrogel depends on the hydrophilicity of the monomer, eg, cured poly(HEMA) absorbs 60° of its weight of water and thus forms a hydrogel with about 38° water content. Other hydrophilic monomers, such as *N*-vinylpyrrolidinone [88-12-0] (NVP) (6), and glycerol methacrylate [100-92-5] (GM) (7), and acrylamide monomers, such as diacetone acrylamide [2873-94-9] (DAA), have also been used to form hydrogels with higher water content.



These monomers impart hydrophilicity as a result of the presence of polar –OH and –CONH– groups. Acid-containing monomers, such as methacrylic acid [79-41-4] (MAA), and 2-acrylamido-2-methylpropanesulfonic acid [15214-89-8], provide ionic character at pH above 7.0 and contribute a large amount of water absorption.

Precursors of hydrophilic monomers, such as vinyl esters and acrylonitrile [107-13-1] have been investigated for hydrogel lens applications. Vinyl esters, eg, vinyl acetate [108-05-4], form poly(vinyl alcohol) [9002-89-5], and absorb substantial amounts of water; cured poly(vinyl alcohol) hydrogels have been claimed useful as contact lenses (114–117). Polyacrylonitrile [25014-41-9] has been hydrolyzed under different controlled conditions to form high water content hydrogel lenses with good mechanical strength resulting from hard segment domain formation during polymerization (118,119). Research efforts on new hydrophilic monomers for high water hydrogel lenses continue into the 1990s.

Hydrophilic monomers must be cured with cross-linkers to obtain dimensionally stable hydrogels. When a methacrylate-based hydrophilic monomer, such as HEMA, is used to form hydrogel, ethylene glycol dimethacrylate [97-90-5] (EGDMA), or poly(ethylene glycol) dimethacrylates of different molecular weights are commonly used as cross-linkers. Trimethacrylates, such as trimethylolpropane trimethacrylate [3290-92-4] (TMPTMA), have also been used. When a hydrogel is formed from a methacrylate monomer and NVP, allyl methacrylate [96-65-9] (AMA) can be used as a cross-linker (120). Cross-linkers provide the required physical shapes and control some of the key mechanical properties, such as modulus and tear strength, of the hydrogels. Thus the choice and amount of a cross-linker are important in the formulation of contact lenses.

Mechanical properties of a hydrogel lens also are affected by the use of a hydrophobic monomer, such as a low alkyl methacrylate. This is particularly important when the water content of the hydrogel lens is very high. The use of these methacrylates helps preserve the required mechanical strength. Methyl methacrylate [80-62-6] (MMA) (121), isobutyl methacrylate [97-86-9] (122), and *n*-pentyl methacrylate [2849-98-1] (123) all have been used for this purpose.

Hydrogel lenses are obtained by photo or thermal polymerization of monomers in bulk or in solution. A variety of thermal and photo catalysts have been used in hydrogel formulations.

Because of the many choices of hydrophilic monomers, cross-linkers, and hydrophobic monomers, a large number of formulations have been developed and manufactured into hydrogel lenses. The water content of these hydrogel lenses ranges from about 38° , for HEMA-based lenses, to 80° , for poly(vinyl alcohol) and partially hydrolyzed acrylonitrile lenses. Table 2 gives a representative list of FDA approved hydrogel materials available to the consumer in the early 1990s.

Table 2. Key Properties of Select Hydrogel Lens

USAN name ^a	CAS Registry Number ^b	Composition	Water content, °	Dk, barrer ^c	Modulus, MPa ^d	Trade name	Manufacturer
bufilcon A ^e	[56030-52-5]	HEMA, DAA, MAA	55	17.0	0.37	Hydrocurve	Sola Barnes-Hind
crofilcon A ^f	[50450-03-8]	GM, MMA	39	8.5	1.52	CSI	Sola Barnes-Hind
etafilcon A ^e		HEMA, MAA, TMPTMA	58	24.0	0.20	Vistamarc	Vistakon
hefilcon A		HEMA, NVP, EGDMA				Acuvue ^g	Vistakon
hefilcon B ^f	[36425-29-3]	HEMA, NVP, EGDMA	45	13.0	0.95	Softsite	Sola Barnes-Hind
lidofilcon A ^h		MMA, NVP, AMA	74	33.0	0.34	Sauflon	Allergan
phemfilcon A ^e		HEMA, EEM ⁱ , MAA	55	17.0	0.27	B&L 70	Bausch & Lomb
polymacon ^f	[25053-81-0]	HEMA, EGDMA	38	9.5	0.56	Durasoft	Wesley-Jessen
						Hydron	Allergan
						Softlens	Bausch & Lomb
						SeeQuence ^g	Bausch & Lomb
tetrafilcon A ^f			43	13.0		AOSoft	Ciba Vision

		HEMA, NVP, MMA, DVB ¹			Aquaflex	Cooper Vision
vifilcon A ^e	[35528-20-2]		55	17.0	Softcon	Ciba Vision
		HEMA, MAA, NVP, EGDMA			NewVue ^g	Ciba Vision

^a U.S. Adapted Name Council designation of the polymeric material.

^b Provided if available.

^c 1 barrer = 10⁻¹¹ cm³ O₂ (at STP)·cm s·cm⁻²·mm Hg.

^d To convert MPa to g mm², multiply by 102.

^e FDA group IV; high protein uptake.

^f FDA group I; low protein uptake.

^g As disposable lens.

^h FDA group II; low protein uptake.

¹ 2-Ethoxyethyl methacrylate [999-61-1].

¹ Divinylbenzene [1321-74-6].

The data in Table 2 indicate that oxygen permeability depends exclusively on water content. To increase oxygen transport to the cornea through a hydrogel lens, the lenses should either be made thinner or be fabricated from hydrophilic monomer with higher hydrophilicity, ie, HEMA-based lenses were fabricated with a lens center thickness of 70 μm for daily wear and 35 μm for extended wear. Some hydrogel lenses, such as etafilcon, are manufactured from formulations containing an acid monomer, such as methacrylic acid, to gain higher water content. Unfortunately, the presence of methacrylate anion invites large amounts of lysozyme uptake, which reduces the water content of the lens and may affect clinical performance.

Silicone Hydrogels. Based on the success in RGP applications, polysiloxanylalkyl acrylates–methacrylates, such as TRIS, have been employed in the preparations of silicone hydrogels. These materials were used in combination with hydrophilic monomers such as HEMA and *N,N*-dimethylacrylamide [2680-03-7] to form hydrogel lenses with high oxygen permeability (16,124–126). They also have been used with methacrylate-based prepolymers, such as those with polyurethane linkages (127,129), to give hydrogel lenses with high oxygen permeability as well as excellent mechanical properties (128,129).

Silicone-based monomers were prepared in different prepolymer forms, such as dimethacrylate-capped polysiloxane (4), and were extensively investigated for hydrogel lens applications. For example, the prepolymer of structure (4) formed hydrogels when copolymerized with a hydrophilic monomer such as *N,N*-dimethylacrylamide or acrylic acid (104,130–132). Polysiloxane prepolymers can be modified to enhance wettability and physical properties of the lenses, ie, methyl groups on the siloxane units in the prepolymers, modified with hydrophilic groups, improve wettability (133,134). The polysiloxane structures can be linked to urethane linkages before end-capping with methacrylate groups (127,128,135–138). When used in hydrogels, these modified prepolymers show better mechanical properties than those without modification. Polysiloxanes also can be end-capped with styrene moieties to give prepolymers that form hydrogels when copolymerized with hydrophilic monomers (139).

In addition to the silicone-based monomers–prepolymers described here, most silicone-based monomers–prepolymers claimed useful as hard lens material have been evaluated for hydrogel lens application. A majority of the efforts in preparing silicone-based hydrogels has been for increased oxygen permeability. Although silicone hydrogel lenses were claimed wettable because of moderate water contents, they have not always proven wettable in clinical testing. A possible explanation for the nonwetting is the migration of siloxane fragments to the hydrophobic air surface. In addition, lipidlike deposits are a potential problem in silicone hydrogel lenses.

Fluorohydrogels. In addition to applications in hard lenses, fluorine-based monomers, such as fluoroalkyl acrylates–methacrylates (140,141) and fluorostyrene and fluorosulfonamide monomers (142), give hydrogel lenses after copolymerizing with hydrophilic monomers such as substituted acrylamide and NVP. In general, these hydrogels claim to have good oxygen permeability and to be useful as extended wear contact lenses (140,141).

Miscellaneous Hydrogel Classes. Because of unique structure and flexibility in manipulating the mechanical properties, polyurethanes have been evaluated for contact lens applications. Nonsilicone polyurethane hydrogels with water contents as high as 20% were claimed to have contact lens application (143). Polyurethane interpenetrating network compositions, formed by polymerizing diacrylates in the presence of a hydrophilic polyurethane, form hydrogels and are useful as contact lenses (144–147). Prepolymers with urethane linkages or well-defined hard–soft–hard polyurethane elastomer blocks are useful in forming hydrogels with improved mechanical properties as well as good oxygen permeability (127,128,135–138,148,149).

There is great interest in developing soft lens materials that have high oxygen permeability and good wettability. Such lenses will provide patients with good comfort, high deposit resistance, and superior corneal health, particularly for extended wear applications.

The popularity of the disposable lens concept is expected to continue, and the future should bring about lenses having enhanced clinical performance combined with the convenience of disposability.

Flexible Nonhydrogel Lenses

Silicone–Fluorosilicone Lenses. Silicone rubber has long been considered a unique contact lens material (55), and the development of silicone rubber lenses has been reviewed in earlier editions of the *Encyclopedia*. The oxygen permeability of silicone rubber, >300 barrers, is virtually unsurpassed by any other polymeric material considered for contact lens applications.

The principal problems for silicone rubber as a viable lens material are the nonpolar nature, which gives lipid deposits and wettability problems; and the tendency to adhere to the cornea. Efforts to modify the silicone lens surface for improved wettability have achieved limited success. These efforts include grafting hydrophilic monomers, such as HEMA, GM (150), and NVP (151–153), to the lens surface and plasma treatments of finished lenses. Efforts to improve the movement of silicone lenses on the cornea with various lens designs have not been successful, and the cause of lens–cornea adherence, which is not an exclusive problem of silicone lenses, is an active area of research.

Other polysiloxane-containing polymerizable systems, such as prepolymers described earlier, also claim to have application as flexible nonhydrogel lenses (129–139,148,149,154). Prepolymers containing block copolymers of polysiloxane and poly(alkylene oxide) give highly oxygen-permeable flexible lenses with good wettability (155–157). These compositions usually have some hydrophilic monomers, particularly an acid monomer, in the composition to enhance wettability. Monomers with fluoroalkyl groups, or polysiloxane prepolymers with fluoroalkyl modified siloxane side chains (158,159), are used in addition to a wetting monomer to minimize lipidlike deposits.

Nonsilicone Flexible Lenses. To avoid wettability and lipid deposit problems associated with silicone, nonsilicone-containing elastomeric materials have been investigated as lens materials. The compositions studied have shown good oxygen permeability. For example, fluorinated telechelic polyether-based prepolymers, derived from perfluorinated poly(ethylene glycol) [25322-68-3], give lenses with good oxygen permeability and wettability after copolymerizing with some hydrophilic monomers (103,160). These lenses, manufactured by 3M Corp., gave low overnight corneal swelling in clinical testing (161). They have been marketed as the Advent lens by Allergan, and have had limited success, because of poor comfort. Prepolymers derived from poly(alkylene oxide)s also give nonhydrogel flexible lenses with reasonable oxygen permeability as high as 70 barrers and good wettability (162–164).

Tinted Lenses

Contact lenses are tinted for cosmetic reasons, ie, to modify or change eye color, and for visibility, ie, to locate the lens more easily when it is out of the eye.

Both rigid and hydrogel lenses can be visibility tinted. However, only hydrogel lenses are cosmetically tinted, because the rigid lens is rarely fit to cover the entire colored portion of the eye. Cosmetic tints were introduced in the early 1980s as an added feature to soft contact lenses (165). A wearer's natural eye color was combined with the tint in the lens to either enhance the existing color, ie, blue on blue, or to change the eye color, ie, a yellow-tinted lens over blue eyes results in green eyes. Tints are sufficiently dark to alter eye color, but not so dark as to affect visible light transmission adversely. Addition of tints reduces the light transmission from >95% for a clear lens to 85° o for the darkest tint (166) and only slightly affects color perception (167,168). Lenses are tinted after lens manufacture with a masking technique to provide a clear annulus around the lens periphery. Because a standard soft contact lens extends past the limbal junction and the iris, the clear annulus was required to prevent the tinted portion from being visible against the white of the eye.

In the late 1980s, a new type of cosmetic-tinted lens appeared on the market. Called an opaque tint, the lens blocks the natural color of the iris and provides a totally new color to the eye (169,170). Brown-eyed individuals, unable to change their eye color with other tinted lenses, can now have any color eye. Doctors have expressed some concern for emmetropic wearers (require no vision correction) because the lenses are then treated as cosmetic and not therapeutic devices; cases of poor lens and eye care are cited (171,172). The opaque-tinted lenses are standard hydrophilic lenses that have a paste of monomer, colorant, and TiO₂ printed on the anterior surfaces in a particular pattern. A small amount of natural eye color and pattern show through to provide a realistic look. Because the tinted portion of the lenses is opaque, the optical zone of the lens must remain free of the printed pattern to prevent visibility problems. Fit and centering of these lenses are critical to provide clear vision. An alternative method of manufacture is to cast a lens containing a clear layer and TiO₂-laden layer. With proper casting and lathing of the TiO₂ layer, a lens with a clear optical zone and clear periphery surrounding an opaque ring is produced. The colorant is added to the front surface to provide the new color. Even unnatural eye shades such as violet or fluorescent colors (173) can be obtained.

Visibility tints are similar to cosmetic tints except that they are significantly lighter in intensity. Visible light transmission losses are typically 1–2° o. With these tints, the lens is visible in the lens package and against a variety of surfaces, and the lens is easier to handle. The tint can be added as cosmetic tints, with a masking technique to produce a clear annulus, or the whole lens can be tinted.

Lens Colorants. Colorants used with rigid lenses are generally pigments cast into the solid matrix during polymerization. The vast majority of tints used for soft contact lenses use dyes originally developed for the textile industry, and classified according to the American Association of Textile Chemists and Colorists (AATCC). Vat, pigment, and reactive dyes are routinely used (see DYES, REACTIVE). Dyes approved by the FDA for use in contact lenses are listed in Table 3 (174) (see also COLORANTS FOR FOODS, DRUGS, AND MEDICAL DEVICES).

Table 3. Dyes Approved by the FDA for Use in Contact Lenses

Color Index name	FDA names	CAS Registry Number	Color Index number
Pigment Violet #23		[6358-30-1]	51319
Pigment Blue #36	chromium–cobalt–aluminum oxide	[68187-11-1]	77343
Pigment Green #17	chromic oxide (green)	[1308-38-9]	77288
Pigment Green #7	phthalocyanine green	[1328-53-6]	74260
Iron Oxide		[977053-38-5]	77491
Pigment White #6	titanium dioxide	[13463-67-7]	77891
Vat Orange #1		[1324-11-4]	59105
Vat Brown #1		[2475-33-4]	70800
Vat Yellow #3		[82-18-8]	61725
Vat Blue #6	D&C Blue #9	[130-20-1]	69825
Vat Green #1		[128-58-5]	59825
Reactive Black #5		[17095-24-8]	20505
Reactive Blue #21		[73049-92-0]	
Reactive Orange #72		[68189-39-9]	17754
Reactive Yellow #15		[60958-41-0]	
Reactive Blue #19		[2580-78-1]	61200
Reactive Blue #4		[4499-01-8]	61205
Reactive Red #11		[12226-08-3]	
Reactive Yellow #86		[61951-86-8]	
Reactive Blue #163		[72847-56-4]	
Reactive Blue #246		[121888-69-5]	61548
Solvent Green #3	D&C Green #6	[128-80-3]	61565
Solvent Violet #13	D&C Violet #2	[81-48-1]	60725
Acid Yellow #3	D&C Yellow #10	[8004-92-0]	47005

Vat dyes are water-soluble in the reduced state but extremely insoluble in the oxidized anthraquinone state (see DYES, ANTHRAQUINONE). As an example, Figure 2 shows the soluble and oxidized states for Vat Yellow #3, used as a yellow colorant in several lenses. The water-soluble form of the dye is allowed to diffuse into the polymer matrix for a specified amount of time. The dye is then treated with a mild oxidizing agent and the insoluble form is precipitated. The tints are extremely colorfast and stable for the life of the lens; however, the dyes are unstable in the soluble form and quite expensive.

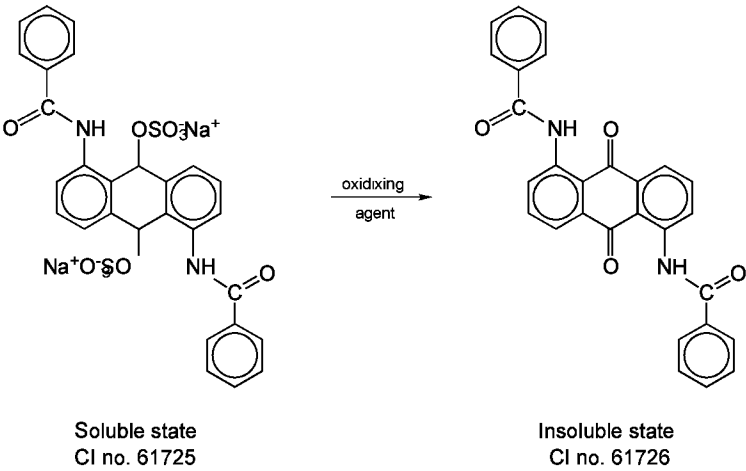


Fig. 2. Reaction of Vat Yellow #3 in contact lens manufacture.

Reactive dyes generally contain colored anthraquinones with various reactive groups attached to the nonchromophoric portions of the molecule. The reactive dyes approved by the FDA are either dichloro-*s*-triazines or sulfatoethyl sulfones, termed vinyl sulfones. Both groups react with pendent hydroxyl groups on the polymer matrix.

A recent addition to Table 3, Reactive Blue 246 differs from the other dyes and is not added to a finished lens. The dye molecule has methacrylate groups attached to an anthraquinone and is incorporated directly into the polymer matrix during polymer cure (175). This in-monomer concept has the potential to reduce dramatically the cost of visibility tinting of a contact lens.

Surface-Modified Lenses

The surface of a contact lens, whether rigid or soft, plays a significant role in the biocompatibility of a particular material. Wettability in the ocular environment, protein and lipid deposition, frictional comfort, and other properties all have significant surface components. Often, a polymer meeting a majority of the requirements for a contact lens application will show deficiencies at the surface. For example, the silicone lens, although possessing excellent oxygen permeability, shows poor wetting in the ocular environment and must be plasma treated to render the surface wettable.

Surface modification of a contact lens can be grouped into physical and chemical types of treatment. Physical treatments include plasma treatments with water vapor (silicone lens) and oxygen (176) and plasma polymerization for which the material surface is exposed to the plasma in the presence of a reactive monomer (177). Surfaces are also altered with exposure to uv radiation (178) or bombardment with oxides of nitrogen (179). Ion implantation (qv) of RGP plastics (180) can greatly increase the surface hardness and hence the scratch resistance without seriously affecting the transmission of light.

Chemical treatments provide a wider variety of applications, depending on the chemical functionality at the original surface and the type of functionality desired as the replacement. Most treatments replace hydrophobic and nonwetting surfaces with hydrophilic and wetting surfaces. Chemical modifications by acylation (181), esterification (182), hydroxylation (183), and treatment with organic anhydrides (184) and alkali (185) have been reported. Isocyanates have also been used as grafting links for hydrophilic coatings (186). Hydrophilic polymers can be grafted onto generally siloxane surfaces to produce a dramatic increase in wettability (187,189).

A good example of a surface-modified lens is the Sola Barnes-Hind Hydrocurve Elite lens, introduced in 1986. The material for the commercial Hydrocurve lens, bufilcon A [56030-52-5], contains methacrylic acid and has a high affinity for protein and subsequent deposition. The surface of the Elite lens was chemically modified with the addition of diazomethane (190) to reduce the surface charge. *In vitro* testing demonstrated a decrease in protein adsorption (191).

Although surface treatments, both physical and chemical, have demonstrated the ability to alter specific properties of contact lens surfaces, most treatments fail as a result of alteration of bulk lens properties, instability of surface treatment, or poor ocular compatibility. Research is expected to continue in the characterization and modification of contact lens surfaces.

Manufacture of Contact Lenses

The first usable contact lenses were produced around 1890. Glass was the only optical material available until the 1930s, and the lenses were manufactured by blowing (with or without a mold), grinding, or a combination of these processes. The radius of curvature of the contact lens surfaces was determined from a cast or impression taken of the front surface of the patient's eye. Around 1936, PMMA (Plexiglas) became available as a contact lens material. Because of its machinability at low temperatures, lathing and polishing of contact lenses with improved accuracy became possible. Molding of lenses from plastic sheets heated to conform to the shape of a cast of a patient's eye, with subsequent lathing of the prescription, also became a method of manufacturing contact lenses. Today, most contact lens materials, including xerogels, ie, hydrogels in their dry state, can be produced in a disklike form (button) that allows lathe cutting for accurate manufacture of contact lenses.

The first hydrogel poly(HEMA) lenses were manufactured using the technique of spin casting (192). The liquid monomer (containing cross-linking agents) is injected into the concave cavity of a spinning plastic mold. The centrifugal force causes the monomer to spread over the surface of the mold in a predictable, controllable manner for the particular prescription of lens being made. The formed liquid lens is then polymerized in the mold via photo curing. The resulting hardened lens can be polished or edged and is either hydrated in the mold and then released as a soft lens or released first and then hydrated. The lens undergoes extraction in hot distilled water to remove any unpolymerized monomer. The finished lens is equilibrated in normal saline (0.9% NaCl), measured, inspected, autoclaved, and packaged before shipment. Thus a spin-cast lens has its front surface (convex) formed by a mold and its back surface (concave) formed by free spinning.

Cast molding is an increasingly used manufacturing process for both rigid gas-permeable and hydrogel contact lenses. In this process, two molds, made from a variety of plastics, are used. A female mold forms the lens front surface (convex) and a male mold forms the lens back surface (concave). The plastic molds are made from metal tools or dies that are usually stainless steel, precision lathed, and polished to the specified lens design. A variety of mold materials are used. The polymerized, hardened lens is released from the mated molds and is processed in much the same way as the spin-cast lenses described above.

Numerous variations to these manufacturing techniques exist. Some contact lenses are produced using combinations, eg, by spin casting or molding one surface and lathing the other. Other hydrogel lenses are polymerized in the mold in the hydrated state rather than in the dry state. Some lenses, eg, silicone elastomer, are molded against metal molds instead of plastic molds, similar to the reaction injection molding process. Factors relating to the physical characteristics of the lens and to the lens cost determine the most appropriate manufacturing technique.

BIBLIOGRAPHY

"Contact Lenses," in *ECT* 3rd ed., Vol. 6, pp. 720–742, by M. F. Refojo, Harvard Medical School.

1. R. B. Mandell, *Contact Lens Practice*, 4th ed., Charles C Thomas, Springfield, Ill., 1988.
2. E. S. Bennett and B. A. Weissman, eds., *Clinical Contact Lens Practice*, J. B. Lippincott, Philadelphia, Pa., 1991.
3. C. Scott, *Contemp. Optom.* **9**(1), 24–32 (1990).
4. J. B. Gelvin, *J. Am. Optom. Assoc.* **61**, 842–847 (1990).
5. B. A. Holden and G. W. Mertz, *Invest. Ophthalmol. Vis. Sci.* **25**, 1161–1167 (1984).
6. B. A. Holden and co-workers, *Optom. Vis. Sci.* **67**, 476–481 (1990).
7. R. C. Tripathi and co-workers, *Int. Ophthalmol. Clin.* **31**(2), 91–120 (1991).
8. B. A. Holden, *Optom. Vis. Sci.* **66**, 717–733 (1989).
9. M. Kotow, T. Grant, and B. A. Holden, *Int. Contact Lens Clin.* **14**(3), 95–99 (1987).
10. S. T. Lin and co-workers, *CLAO J.* **17**(1), 44–50 (1991).
11. A. Tomlinson, *Contact Lens Spectrum*, 27–35 (Sept. 1990).
12. J. H. B. Ang and N. Efron, *Int. Contact Lens Clin.* **16**, 48–58 (1989).
13. American Society for Testing and Materials, *Annual Book of ASTM Standards*, Vol. 8.01, ASTM, Philadelphia, Pa., 1991, D1434.
14. Y. Mizutani, H. Iwashita, S. Nozaki, T. Harata, and Y. Mizutani, *J. Jpn. Contact Lens Soc.* **30**, 318–325 (1988).
15. S. Aiba, M. Ohashi, and S. Y. Huang, *Ind. Eng. Chem. Fundam.* **7**, 497–502 (1968).
16. I. Fatt and R. St. Helen, *Am. J. Optom.* **48**, 545–555 (1971).
17. B. A. Weissman and I. Fatt, *Optom. Vis. Sci.* **66**, 235–238 (1989).
18. I. Fatt, J. E. Rasson, and J. B. Melpolder, *Int. Contact Lens Clin.* **14**, 389–402 (1987).

19. N. A. Brennan, N. Efron, and S. D. Newman, *Int. Contact Lens Clin.* **14**, 407–411 (1987).

20. I. Fatt and J. Morris, *Optician* **174**, 17–20 (1977).

21. M. H. Johnson, R. L. Boltz, and L. B. Godio, *Int. Eyecare* **2**, 423–431 (1986).

22. *Tyler's Quarterly*, Vol. 8, Tyler's Quarterly Publications, Little Rock, Ark., June, 1991.

23. G. Y. Mousa, M. G. Callender, J. G. Sivak, and D. J. Egan, *Int. Contact Lens Clin.* **10**, 31–37 (1983).

24. Ref. 13, D542.

25. G. N. Orsborn and S. G. Zantos, *CLAO J.* **14**, 81–85 (1988).

26. N. A. Brennan, R. Lowe, N. Efron, J. L. Ungerer, and L. G. Carney, *Am. J. Optom. Phys. Optics* **64**, 534–539 (1987).

27. F. X. Quinn, E. Kampff, G. Smyth, and V. J. McBrierty, *Macromolecules* **21**, 3191–3198 (1988).

28. F. X. Quinn, V. J. McBrierty, A. C. Wilson, and G. D. Friends, *Macromolecules* **23**, 4576–4581 (1990).

29. D. W. Larsen, J. W. Huff, and B. A. Holden, *Curr. Eye Res.* **9**, 697–706 (1990).

30. Ref. 13, D1708.

31. R. W. Stevenson, *Optom. Vis. Sci.* **68**, 142–145 (1991).

32. Ref. 13, D1938.

33. I. Fatt, *Int. Contact Lens Clin.* **14**, 360–367 (1987).

34. A. S. Bruce, and S. J. Dain, *Int. Contact Lens Clin.* **15**, 276–281 (1988).

35. M. R. Lattimore, *Optom. Vis. Sci.* **66**, 117–122 (1989).

36. T. W. Bochow, S. K. West, A. Azar, B. Munoz, A. Sommer, and H. R. Taylor, *Arch. Ophthalmol.* **107**, 369–372 (1989).

37. J. P. G. Bermanson, D. G. Pitts, and L. W. F. Chu, *J. Am. Optom. Assoc.* **59**, 178–182 (1988).

38. B. R. Chou, A. P. Cullen, and K. A. Dumbleton, *Int. Contact Lens Clin.* **15**, 244–251 (1988).

39. M. D. Sarver, L. Bowman, R. B. DiMartino, D. Lau, and W. Umeda, *Int. Contact Lens Clin.* **11**, 479–488 (1984).

40. L. Wilhelmy, *Ann. Physik.* **119**, 177 (1863).

41. D. E. Gregonis, R. Hsu, D. E. Buerger, L. M. Smith, and J. D. Andrade in R. B. Seymour and G. A. Stahl, eds., *Macromolecular Solutions*, Pergamon Press, New York, 1982.

42. W. J. Benjamin, D. Y. Yeager, N. N. Desai, and C. A. Carmichael, *Int. Eyecare* **2**, 163–170 (1986).

43. D. Briggs, H. Chan, M. J. Hearn, D. I. McBriar, and H. S. Munro, *Langmuir* **6**, 420–424 (1990).

44. L. M. Smith, L. Bowman, and J. D. Andrade, *Proceedings of the Durham, England Conference on Biomedical Polymers*, July 12–15, 1982.

45. D. Mirejovsky, A. S. Patel, and D. D. Rodriguez, *Curr. Eye Res.* **10**, 187–196 (1991).

46. S. Stenson, *Arch. Ophthalmol.* **101**, 402–404 (1983).

47. J. Mizutani, Y. Mizutani, and T. Ito, *Contact Lens J.* **16**, 3–12 (1988).

48. P. J. Caroline and co-workers, *CLAO J.* **11**, 311–316 (1985).

49. D. R. Absolom, W. Zingg, and A. W. Neumann, *J. Biomed. Mater. Res.* **21**, 161–171 (1987).

50. W. Norde, *Adv. Colloid Interface Sci.* **25**, 267–288 (1986).

51. J. L. Bohnert, T. A. Horbett, B. D. Ratner, and F. H. Royce, *Invest. Ophthalmol. Vis. Sci.* **29**, 362–372 (1988).

52. C. D. Leahy, R. B. Mandell, and S. T. Lin, *Optom. Vis. Sci.* **67**, 504–511 (1990).

53. N. O. Stahl, L. A. Reich, and E. Ivani, *J. Am. Optom. Assoc.* **45**, 302 (1974).

54. **U.S. Pat. 3,900,250** (Aug. 19, 1975), E. J. Evani (to Rynco Scientific Corp.).

55. **U.S. Pat. 3,341,490** (July 12, 1967), D. F. Burdick, J. L. Mishler, and K. E. Polmanteer (to Dow Corning).

56. **U.S. Pat. 3,808,178** (Apr. 30, 1974), N. G. Gaylord (to Polycon Labs).

57. **U.S. Pat. 4,120,570** (Oct. 17, 1978), N. G. Gaylord (to Syntex Inc.).

58. **U.S. Pat. 4,216,303** (Aug. 5, 1980), N. N. Novisky.

59. **U.S. Pat. 4,152,508** (May 1, 1979), E. J. Ellis and J. C. Salamone (to Polymer Technology Corp.).

60. **U.S. Pat. 4,661,573** (Apr. 28, 1987), D. J. Ratkowski and W. J. Burke (to Paragon Optical Inc.).

61. **U.S. Pat. 4,812,598** (Mar. 14, 1989), R. C. S. Chen (to Ocular Technologies, Inc.).

62. **U.S. Pat. 4,645,811** (Feb. 24, 1987), J. J. Falcetta and W. F. Kunzler (to Oculus Contact Lens Co.).

63. **U.S. Pat. 4,424,328** (Jan. 3, 1984), E. J. Ellis (to Polymer Technology Corp.).

64. **U.S. Pat. 4,604,479** (Aug. 5, 1986), E. J. Ellis (to Polymer Technology Corp.).

65. **U.S. Pat. 4,766,189** (Aug. 23, 1988), G. F. Tsuetaki (to Fused Kontacts of Chicago).

66. **Can. Pat. 1,205,247** (May 27, 1986), A. R. LeBoeuf and H. K. Singer (to American Optical Corp.).

67. **U.S. Pat. 4,242,483** (Dec. 10, 1980), N. N. Novisky.

68. **U.S. Pat. 4,248,989** (Feb. 3, 1981), N. N. Novisky.

69. **U.S. Pat. 4,780,516** (Oct. 25, 1988), W. M. Foley.

70. **U.S. Pat. 4,886,864** (Dec. 12, 1989), W. M. Foley.

71. **U.S. Pat. 4,711,943** (Dec. 8, 1987), T. B. Harvey (to Sola U. S. A. Inc.).

72. **U.S. Pat. 4,649,185** (Mar. 10, 1987), M. Takamizawa (to Shin-Etsu Chemical Co.).

73. **Eur. Pat. A2 0 220 919** (June 5, 1987), M. Takamizawa (to Shin-Etsu Chemical Co.).

74. **U.S. Pat. 4,923,906** (May 8, 1990), K. F. Mueller and P. Harisiades (to CIBA-GEIGY Corp.).

75. **Eur. Pat. 396,364** (Nov. 7, 1990), R. E. Bambury and D. E. Seelye (to Bausch & Lomb).

76. **U.S. Pat. 4,594,401** (June 10, 1986), S. Ichinohe, K. Isobe, Y. Tanaka, and K. Kakahashi (to Shinetsu Chemical Co.).

77. **U.S. Pat. 4,611,039** (Sept. 9, 1986), J. C. Powell and G. G. Allan (to Precision-Cosmet Co.).

78. T. Masuda, E. Isobe, T. Higashimura, and K. Takada, *J. Am. Chem. Soc.* **105**, 7473 (1983).

79. T. Masuda, E. Isobe, and T. Higashimura, *Macromolecules* **18**, 121 (1987).

80. C. F. Tien, A. C. Savoca, A. D. Sumamer, and M. Langsam, *ACS Polym. Mater. Sci. Eng. Proc.* **61**, 507 (1989).

81. T. Otsu, O. Ito, and N. Toyoda, *Makromol. Chem. Rapid Commun.* **2**, 725 (1981).

82. N. Toyoda and T. Otsu, *J. Macromol. Sci. Chem.* **A19**(7), 1011 (1983).

83. T. Otsu, T. Yasuhara, K. Shiraishi, and S. Mori, *Polym. Bull.* **12**, 449 (1984).

84. T. Otsu, A. Matsumoto, and K. Fukushima, *J. Chem. Soc., Chem. Commun.*, 1766 (1985).

85. Y. Murata and J. Hirano, *Chem. Econ. Eng. Rev.* **10**, 18 (1985).

86. **Jpn. Pat. 87 39,589** (Feb. 20, 1987), A. Kubota and T. Mogami (to Seiko Epson K. K.).

87. **Jpn. Pat. 87 50,728** (Mar. 5, 1987), T. Mogami and A. Kubota (to Seiko Epson K. K.).

88. **U.S. Pat. 4,868,260** (Sept. 19, 1989), T. Kawaguchi (to Tomei Sangyo Kabushiki Kaisha).

89. **Jpn. Pat. 90 220,023** (Sept. 3, 1990), T. Otsu and co-workers (to Nippon Oil and Fats Co., Seiko Epson Corp.).

90. **Jpn. Pat. 90 220,022** (Sept. 3, 1990), T. Otsu and co-workers (to Nippon Oil and Fats Co., Seiko Epson Corp.).

91. **U.S. Pat. 4,933,406** (June 12, 1990), N. Amaya, K. Amami, and M. Egwa (to Nippon Oil and Fats Co.).

92. **Jpn. Pat. 90 292,816** (Dec. 5, 1990), K. Anami, N. Amaya, T. Murata, and T. Miyazaki (to Nippon Oil and Fats Co.).

93. **Jpn. Pat. 90 292,817** (Dec. 5, 1990), K. Anami, N. Amaya, T. Murata, and T. Miyazaki (to Nippon Oil and Fats Co.).

94. **U.S. Pat. 3,808,179** (Apr. 30, 1974), N. G. Gaylord (to Polycon Labs).

95. **Can. Pat. 1,205,943** (June 10, 1986), K. Kawamura, S. Yamashita, Y. Yokiyama, and M. Tsuchiya (to Hoya Lens Corp.).

96. **U.S. Pat. 4,665,144** (May 12, 1987), A. Ohmori and N. Tomihashi (to Daikin Industries Inc.).

97. **U.S. Pat. 4,861,850** (Aug. 29, 1989), N. N. Novicky (to Progressive Chemical Research, Ltd.).

98. **Eur. Pat. A2 0 213 412** (Mar. 11, 1987), T. Suzuki, M. Ikari, S. Salto, A. Ohmori, and T. Yasuhara (to Kuraray Co. Ltd.).

99. **Eur. Pat. A2 0 219 312** (Apr. 22, 1987), E. J. Ellis and J. Y. Ellis (to Polymer Technology Co.).

100. **Eur. Pat. A2 1 231 572** (Aug. 12, 1987), N. N. Novisky (to Progressive Chemical Research, Ltd.).

101. **Eur. Pat. A2 0 340 636** (Apr. 30, 1988), T. Sawamoto, M. Nomura, and N. Tarumi (to Hoya Corp.).

102. **U.S. Pat. 4,829,137** (May 9, 1989), N. Stoyan.

103. **U.S. Pat. 4,440,918** (Apr. 3, 1984), D. E. Rice and J. V. Ihlenfeld (to 3M Corp.).

104. **U.S. Pat. 4,153,641** (May 8, 1979), W. G. Deichert, K. C. Su, and M. F. Van Buren (to Bausch & Lomb).

105. A. J. Berebblatt, *Contact Lens Forum* **13**, 48 (May 1988).

106. **U.S. Pat. 4,943,150** (July 24, 1990), W. G. Deichert and R. J. Wrue (to Bausch & Lomb).

107. V. Kudela, in J. I. Kroschwitz, ed., *Polymers: Biomaterials and Medical Applications*, Encyclopedia Reprint Series, John Wiley & Sons, Inc., New York, 1989.

108. M. Rohr, in J. F. Herschel, ed., *Die Brille als Instrument*, Vol. 3, 1921; O. Wichterle, D. Lim, and M. Dreifus, *Ceska. Oftalmol* **17**, 70 (1961).

109. O. Wichterle and D. Lim, *Nature*, **185**, 117 (1961).

110. **U.S. Pat. 2,976,576** (Mar. 28, 1961), O. Wichterle and D. Lim.

111. **U.S. Pat. 3,220,960** (Nov. 30, 1965), O. Wichterle and D. Lim.

112. **U.S. Pat. 3,408,429** (Oct. 29, 1968), O. Wichterle (to Czechslovak Academy of Science).

113. G. E. Lowther, *Int. Eyecare* **2**, 298 (1986).

114. **U.S. Pat. 4,619,793** (Oct. 28, 1986), P. I. Lee (to CIBA-GEIGY Corp.).

115. **Can. Pat. 1 222 357** (June 2, 1987), P. I. Lee (to CIBA-GEIGY Corp.).

116. **U.S. Pat. 4,618,649** (Oct. 2, 1986), R. F. Ofstead (to 3M Corp.).

117. **U.S. Pat. 4,693,939** (Sept. 15, 1987), R. F. Ofstead (to 3M Corp.).

118. **U.S. Pat. 4,943,618** (July 24, 1990), V. A. Stoy, G. P. Stoy, and J. Lovy (to Kingston Technology).

119. R. C. Capozza, W. E. Meyers, H. H. Neidlinger, and V. A. Stoy, paper presented at American Chemical Society National Meeting, Aug. 1990 in *Polym. Preprints* **31**(2), 57 (1990).

120. **U.S. Pat. 3,926,892** (Dec. 16, 1975), F. O. Holcombe (to Burton, Parsons & Co. Inc.).

121. M. J. Popovich, *J. Am. Optom. Assoc.* **47**, 305 (1976).

122. **Jpn. Kokai 75 129,648** (Oct. 12, 1975), H. Magatani, H. Atsuzawa, and Y. Kosaka (to Tokyo Contact Lens Institute).

123. **Jpn. Kokai 76 56,893** (May 18, 1976), K. Tanaka (to Tokyo Contact Lens Co. Ltd.).

124. **U.S. Pat. 4,182,822** (Jan. 8, 1980), S. H. Chang.

125. **U.S. Pat. 4,343,927** (Aug. 10, 1982), S. H. Chang.

126. **U.S. Pat. 4,525,563** (June 25, 1985), I. Shoji, I. Kenichi, S. Takanori, and Y. Tpshiharu (to Tokyo Contact Lens Co.).

127. Y. C. Lai and L. J. Baccei, *J. Appl. Polym. Sci.* **42**, 2039 (1991).

128. **U.S. Pat. 5,034,461** (July 23, 1991), Y. C. Lai and L. J. Baccei (to Bausch & Lomb).

129. Y. C. Lai and L. J. Baccei, *J. Appl. Polym. Sci.* **42**, 3173 (1991).

130. **U.S. Pat. 4,254,248** (Mar. 9, 1981), G. D. Friends and M. F. Van Buren (to Bausch & Lomb).

131. **U.S. Pat. 4,276,402** (June 30, 1981), R. C. Chromrcek, W. G. Deichert, J. J. Falcetta, and M. F. Van Buren (to Bausch & Lomb).

132. **U.S. Pat. 4,277,595** (July 7, 1981), W. G. Deichert, G. C. Niu, and M. F. Van Buren (to Bausch & Lomb).

133. **U.S. Pat. 4,259,467** (Mar. 31, 1981), P. L. Keogh, J. F. Kunzler, and G. C. Niu (to Bausch & Lomb).

134. **U.S. Pat. 4,260,725** (Apr. 7, 1981), P. L. Keogh, J. F. Kunzler, and G. C. Niu (to Bausch & Lomb).

135. **U.S. Pat. 4,136,250** (Jan. 23, 1979), K. F. Mueller and E. K. Kleiner (to CIBA-GEIGY Corp.).

136. **U.S. Pat. 4,486,577** (Dec. 4, 1984), K. F. Mueller, S. J. Heiber, and W. L. Plankl (to CIBA-GEIGY Corp.).

137. **U.S. Pat. 4,605,712** (Aug. 12, 1986), K. F. Mueller, D. Lohmann, and R. A. Falk (to CIBA-GEIGY Corp.).

138. **Eur. Pat. 0 401 990 A2** (Dec. 12, 1990), Y. C. Lai and L. J. Baccei (to Bausch & Lomb).

139. **U.S. Pat. 4,640,941** (Feb. 3, 1987), J. J. Falcetta and J. Park (to Alcon Labs).

140. **U.S. Pat. 4,433,111** (Feb. 21, 1984), H. J. Gee and B. J. Tighe (to Kelvins Lenses Lts.).

141. **U.S. Pat. 4,493,910** (Jan. 15, 1985), H. J. Gee and B. J. Tighe (to Kelvins Lenses Lts.).

142. **Eur. Pat. 0 253 515 A2** (Jan. 20, 1988), J. C. Salomone (to Optimers Co.).

143. **U.S. Pat. 3,786,035** (Jan. 15, 1974), E. A. Blair and D. E. Hudgiu (to Frigitronics).

144. **U.S. Pat. 4,359,558** (Nov. 16, 1982), F. E. Gould and C. W. Johnston (to Tyndale Plains-Hunter).

145. **U.S. Pat. 4,408,023** (Oct. 4, 1983), F. E. Gould and C. W. Johnston (to Tyndale Plains-Hunter).

146. **U.S. Pat. 4,454,309** (June 12, 1984), F. E. Gould and C. W. Johnston (to Tyndale Plains-Hunter).

147. **U.S. Pat. 4,496,535** (Jan. 29, 1985), F. E. Gould and C. W. Johnston (to Tyndale Plains-Hunter).

148. **U.S. Pat. 4,487,904** (Dec. 11, 1984), T. Fukuda and S. Sakamoto (to Toray Ind.).

149. **U.S. Pat. 4,886,866** (Dec. 12, 1989), J. A. Braatz and C. L. Kehr (to W. R. Grace).

150. Fr. Demande 2,208,755 (June 28, 1974), E. W. Merrill (to High Voltage Engineering Co.).

151. H. Yasuda and M. F. Refojo, *J. Polym. Sci. Part A 2*, 5093 (1964).

152. **U.S. Pat. 3,700,573** (Oct. 24, 1972), J. Laizier and G. Wajs (to Commissariat a l'Energie Atomique, France).

153. **U.S. Pat. 3,959,102** (May 25, 1976), W. Lenne and G. Wajs (to Essilor International Fr).

154. **U.S. Pat. 4,962,178** (Oct. 9, 1990), P. Harisiades (to CIBA-GEIGY Corp.).

155. **U.S. Pat. 4,740,533** (Apr. 26, 1988), K. C. Su and R. Robertson (to CIBA-GEIGY Corp.).

156. **Eur. Pat. 0 330 616** (Aug. 30, 1989), K. C. Su and R. Robertson (to CIBA-GEIGY Corp.).

157. **Eur. Pat. 0 330 617** (Aug. 30, 1989), K. C. Su and R. Robertson (to CIBA-GEIGY Corp.).

158. **U.S. Pat. 4,810,764** (Mar. 7, 1989), G. D. Friends and J. F. Kunzler (to Bausch & Lomb).

159. **U.S. Pat. 4,954,586** (Sept. 4, 1990), N. Toyoshima and co-workers (to Menicon Co. Ltd.).

160. **U.S. Pat. 4,818,801** (Apr. 4, 1989), D. E. Rice and J. V. Ihienfeld (to 3M Corp.).

161. *Contact Lens J.* **17**(1), 21 (1989).

162. **U.S. Pat. 4,780,488** (Oct. 25, 1988), K. C. Su and F. F. Molock (to CIBA-GEIGY Corp.).

163. **U.S. Pat. 4,857,606** (Aug. 15, 1989), K. C. Su and F. F. Molock (to CIBA-GEIGY Corp.).

164. **Eur. Pat. 0330 614** (Aug. 30, 1989), F. Molock and K. C. Su (to CIBA-GEIGY Corp.).

165. P. L. Rakow, *J. Ophthalmic Nurs.* **5**, 32 (1986).

166. A. S. Bruce, *Am. J. Optom.* **63**, 941 (1986).

167. A. Tan, *Clin. Exp. Optom.* **70**, 78 (1987).

168. M. Laxer, *Int. Contact Lens Clin.* **17**, 88 (1990).

169. **U.S. Pat. 4,460,523** (July 17, 1984), C. Neefe.

170. **U.S. Pat. 4,472,327** (Sept. 18, 1984), C. Neefe.

171.

K. J. Johns, *Am. J. Ophthalmol.* **105**, 210 (1988).

172.

U. Schell, *Optom. Times* **3**, 8 (1985).

173.

U.S. Pat. 4,702,574 (Oct. 27, 1987), R. Bawa (to Bausch & Lomb).

174.

Code of Federal Regulations, Part 74 (1989).

175.

U.S. Pat. 4,997,897 (Mar. 5, 1991), J. Melpolder (to Bausch & Lomb).

176.

U.S. Pat. 4,214,014 (July 22, 1980), P. Hofer and W. Kohl (to Titmus Eurocon Kontaktlinsen KG DE).

177.

H. Yasuda and co-workers, *J. Biomed. Mater. Res.* **9**, 629–643 (1975).

178.

U.S. Pat. 4,312,725 (Jan. 26, 1982), S. Loshaek and C. M. Shen (to Wesley-Jessen Inc.).

179.

U.S. Pat. 4,409,258 (Oct. 11, 1983), P. Baronet and B. Feurer (to Alcon Pharmaceuticals, Ltd.).

180.

U.S. Pat. 4,743,493 (May 10, 1988), R. W. Oliver and P. Sioshansi (to Spire Corp.).

181.

U.S. Pat. 4,925,698 (May 15, 1990), M. Klausner, M. P. Manning, and R. F. Baddour (to MatTek Corp.).

182.

U.S. Pat. 4,666,249 (May 19, 1987), R. E. Bauman, L. E. Bowman, S. K. Chandrasekaran, T. B. Harvey, and S. B. Lin (to Sola USA Inc.).

183.

U.S. Pat. 4,965,026 (Oct. 23, 1990), R. A. Janssen (to CIBA-GEIGY Corp.).

184.

U.S. Pat. 4,687,816 (Aug. 18, 1987), L. E. Bowman, S. K. Chandrasekaran, T. B. Harvey, and S. B. Lin (to Sola USA Inc.).

185.

U.S. Pat. 4,550,001 (Oct. 29, 1985), T. Ito, Y. Kiyomatsu, T. Shimizu, and T. Suminoi (to Japan Synthetic Rubber).

186.

U.S. Pat. 4,280,970 (July 28, 1981), R. E. Kesting (to Puro pore Inc.).

187.

U.S. Pat. 4,761,436 (Aug. 2, 1988), Y. Kohno and co-workers (to Shin-Etsu Chemical Co.).

188.

U.S. Pat. 4,968,532 (Nov. 6, 1990), E. M. Freeman, R. A. Janssen, and E. C. McCraw (to CIBA-GEIGY Corp.).

189.

A. Baszkin, J. E. Prouse, and M. M. Boissonnade, *Biomaterials* **5**, 175–179 (1984).

190.

U.S. Pat. 4,569,858 (Feb. 11, 1986), D. Lim and P. Morris (to Barnes-Hind Inc.)

191.

A. Greco, *Int. Eyecare* **2**, 625–626 (1986).

192.

L. D. Clements, in M. Ruben, ed., *Soft Contact Lenses*, John Wiley & Sons, Inc., New York, 1978.

Yu-chin Lai
Alan C. Wilson
Steve G. Zantos
Bausch & Lomb

CONTAINERS.

See PACKAGING CONTAINERS AND INDUSTRIAL PRODUCTS.

CONTRACEPTIVES

Women have searched for effective methods of birth control since antiquity. From the earliest times, most have considered the ideal method one that would safely avert unwanted pregnancy while also avoiding inconvenience during intercourse. Some of the contraceptive methods practiced historically were based on superstitions and taboos; eg, the ancient Chinese advised women to swallow 24 live tadpoles in early spring, believing that this would ensure conception-free years. Other methods evolved that were more effective, and had a basis in rational chemical concepts. For example, the use of chemical agents placed in the vagina to prevent pregnancy was recorded in ancient Egyptian and Greek writings (1).

Today, millions of people throughout the world use a variety of methods to regulate human fertility, including chemical methods, eg, oral contraceptives, vaginal contraceptives, injectable implantable contraceptives, and contragestational agents; intrauterine devices; intravaginal barrier methods; male and female condoms; surgical sterilization; induced abortion; and natural family planning. Each of these has its advantages and disadvantages, and it is generally accepted that none of these methods represents an ideal method of fertility regulation (2–4) for everyone.

This article reviews various contraceptive methods, with particular emphasis on the evolution of the chemical methods used in current hormonal contraceptives and contragestational products, and describes research efforts directed toward the development of new approaches to control human fertility.

Oral Contraceptives

Oral contraceptives were first introduced in the early 1960s. Since then there have been many changes in these products, including lowering of the contraceptive dose and the introduction of phasic oral contraceptives (5). Much has been written about the impact of oral contraceptives on the practice of family planning (6). The availability of an oral, hormonal contraceptive, ie, the pill, has permitted women to make responsible decisions in matters regarding size of family, spacing of children, and choice of lifestyle. The introduction of oral contraceptives also opened the door to intensified research in basic and applied reproductive biology. This led to the introduction of not only the second and third generation oral contraceptives used today, but also to today's broad range of other methods of fertility regulation.

Oral contraceptives are among the most popular form of reversible contraception in most countries. As of the early 1990s, over 60 million women around the world use the pill and almost 150 million women have used oral contraceptives sometime during their reproductive lives (7). The commercial market for oral contraceptives is large and expanding. U.S. drug store purchases alone will exceed \$1 billion in 1992.

History. Detailed reviews of biological, chemical, and clinical research that led to the introduction of oral contraceptives are available in the scientific and medical literature (8,9).

The concept of hormonal control of ovulation first appeared in 1921 when Ludwig Haberlandt, a physiologist at the University of Innsbruck, showed that extracts of the corpus luteum, containing progesterone [57-83-0], C₂₁H₃₀O₂, could inhibit ovulation and make mice and rabbits infertile (10). Subsequent research showed that this phenomenon occurred in other species as well.

Making progesterone in the laboratory was both difficult and expensive, and progesterone could not be given orally because its natural form is destroyed in the digestive system. During the next 20 years, other sex hormones were isolated, identified, and synthesized. Unfortunately, evaluation of their clinical utility, like that of progesterone, was severely restricted by insufficient quantities of compounds. By 1943, Russell Marker, an American chemist, had developed a method for synthesizing progesterone and other steroid derivatives from diogenin, a naturally occurring steroid present in a species of *Dioscorea*. Marker's studies led to the synthesis of larger quantities of steroids that could be clinically evaluated for their effects on fertility regulation. It was discovered that elimination of the methyl group at C-19 of progesterone results in a progestational compound that is orally active. Also, structural modification of steroids from other hormone classes, such as the elimination of C-19 from testosterone, results in orally active compounds that have progesterone-like effects, ie, they induce secretory changes in estrogen-primed endometrium; induce the formation of thick, viscous cervical mucus; and suppress the release of luteinizing hormone and ultimately ovulation. These compounds are commonly termed progestogens or progestins. In 1951, Djerassi at Syntex and Coulton at Seale in the United States produced substituted 19-nor-synthetic progestogens which are still utilized in oral contraceptives, ie, norethisterone (1) (norethindrone in the United States) and norethynodrel (2), respectively (Fig. 1). Other progestogens which were subsequently synthesized include lynestrenol (5), chlormadinone acetate [302-22-7], C₂₃H₂₉ClO₄, medroxyprogesterone acetate (6), ethynodiol diacetate (4), levonorgestrel (11), desogestrel (8), norgestimate (10), and gestodene (9).

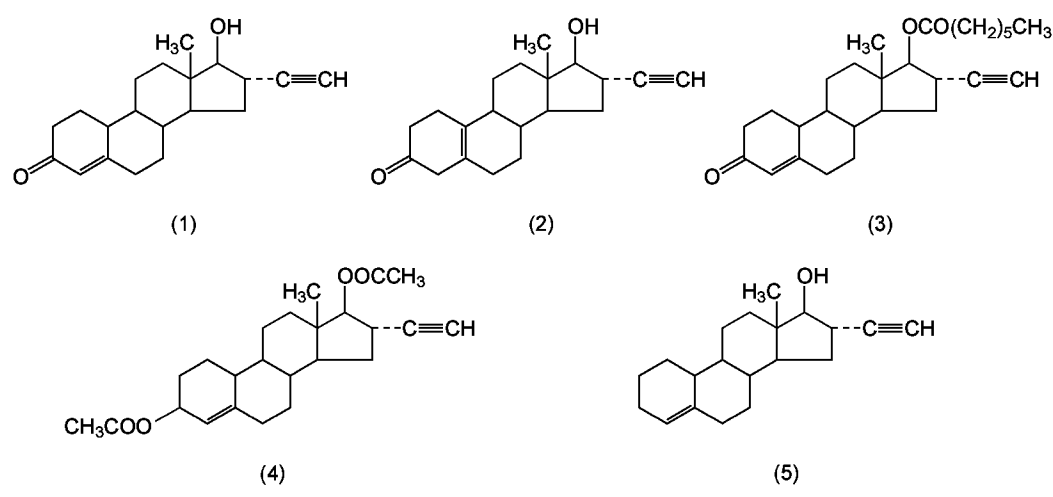


Fig. 1. Examples of estrane progestogens. (1) Norethindrone [68-22-4] (norethisterone, C₂₀H₂ O₂); (2) norethynodrel [68-23-5] C₂₀H₂ O₂; (3)

norethindrone enanthate [3836-23-5], C₂₇H₃₈O₃; (4) ethynodiol diacetate [297-76-7], C₂₄H₃₂O₄; (5) ethinylestrenol [52-76-7] (lynestrenol, C₂₀H₂₈O).

The first clinical study of an oral contraceptive began with a pill investigators believed contained only the progestogen norethynodrel. Early clinical data demonstrated that this preparation inhibited ovulation. In subsequent studies, utilizing newly synthesized batches of the progestogen, ovulation was inhibited, but significant intermenstrual vaginal bleeding also occurred. It was discovered that the original batch of progestogen, which gave better cycle control, was contaminated with an estrogen, mestranol (13). It was subsequently shown that addition of estrogen to pure progestogen also increased efficacy. This realization led to the controlled inclusion of estrogen in oral contraceptives (11).

Second and Third Generation Oral Contraceptives. Most oral contraceptives are combinations of an estrogenic agent and a progestational agent (progestogen). Estrogens are found in the ovary, and are important in preventing pregnancy; they work in conjunction with the progestogen to suppress ovulation.

The estrogenic and progestational components provide their primary contraceptive effect by blocking ovulation, ie, preventing the selection of a dominant follicle in the ovary by a negative feedback action on the hypothalamus and pituitary. This inhibits pituitary secretion of follicle-stimulating hormone (FSH) [9002-18-0] and luteinizing hormone (LH) [9002-67-9], with the resultant inhibition of ovulation. The estrogen component also provides stability to the endometrium so that unwanted breakthrough bleeding can be avoided. This combination also provides several ancillary contraceptive mechanisms by interfering with fertilization and implantation processes should ovulation occur (7).

Early clinical investigators were concerned primarily about the contraceptive efficacy of the first oral contraceptive products. These original oral contraceptives were relatively high dose products as judged by later standards. Epidemiological studies during the 1960s and 1970s indicated that oral contraceptive usage was associated with an increased risk of thromboembolic disease. These cardiovascular complications were associated with the relatively high doses of estrogen used in original oral contraceptives, especially in women who smoke. These findings led to the development of low estrogen dose combination oral contraceptives. Early oral contraception formulations contained up to 100–150 µg of estrogen. By the early 1990s the majority of oral contraceptives contained only 30 or 35 µg of estrogen (Table 1). These low dose products are commonly referred to as second-generation oral contraceptives, ie, new products which have almost totally replaced the original high dose oral contraceptives. Numerous review articles and reports on cohort studies describing the use of oral contraceptives and the incidence of side effects are available (12–22).

Table 1. Low Dose Monophasic Estrogen–Progestogen^a Combination Oral Contraceptives Marketed in the United States

Trade name	Estrogen, µg	Progestogen, mg	Launch date	Manufacturer
Bevicon	35	0.50	10 75	Syntex
Demulen 1 35	35	1.00 ^b	1 82	Seale
Demulen 1 50	50	1.00 ^b	12 70	Seale
Desogen	30	0.15	1 93	Organon
Genora 0.5 35	35	0.50	10 89	Rugby Labs
Genora 1 35	35	1.00	12 86	Rugby Labs
Genora 1 50	50 ^c	1.00	10 86	Rugby Labs
Levlen	30	0.15 ^d	1 86	Berlex
Loestrin 1.5 30	30	1.50 ^e	10 73	Parke-Davis
Loestrin 1 20	20	1.00 ^e	10 73	Parke-Davis
Lo-ovral	30	0.30 ^d	3 75	Wyeth
M.E.E.	35	1.00	3 88	Lexis Pharm
Modicon	35	0.50	12 74	Ortho
Nelova 0.5 35E	35	0.50	11 87	Warner-Chilcott
Nelova 1 35E	35	1.00	11 87	Warner-Chilcott
Nelova 1 50M	50 ^c	1.00	11 88	Warner-Chilcott
Norcept-E 1 35	35	1.00	8 89	Gynopharma
Nordette	30	0.15 ^d	5 82	Wyeth
Norinyl + 35	35	1.00	12 83	Syntex
Norinyl + 50	50 ^c	1.00	12 83	Syntex
OrthoCept	30	0.15	1 93	Ortho
Ortho Novum 1 35	35	1.00	1 80	Ortho
Ortho Novum 1 50	50 ^c	1.00	4 67	Ortho
Ortho-Cyclen	35	0.25 ^f	10 92	Ortho
Ovcon 35	35	0.40	4 76	Mead Johnson
Ovcon 50	50	1.00	9 78	Mead Johnson
Ovral	50	0.50 ^d	9 76	Wyeth

^a Ethinyl estradiol-norethindrone unless otherwise noted.
^b Ethynodiol diacetate (4).
^c Mestranol (13).
^d Norgestrel (11).
^e Norethindrone acetate.
^f Norgestimate.

As companies continued to conduct studies to find the lowest doses of estrogen and progestogen effective as contraceptives, women began to experience increased levels of intermenstrual bleeding. This observation led to the development of multiphasic oral contraceptives, a new approach to low-dose contraception (Table 2). Multiphasic oral contraceptives vary the dose of active ingredients or the ratio of progestogen to estrogen throughout the cycle, instead of remaining constant as in conventional combination oral contraceptives, to utilize the lowest effective dose of active ingredients yet still control intermenstrual bleeding. Some, but not all, of the low dose multiphasic oral regimens studied significantly reduce intermenstrual bleeding and spotting (23). The success of this approach was seen in the marketplace when the first triphasic oral contraceptive was introduced in the United States (Ortho-Novum 7 7 7) in 1984. It is now the most commonly prescribed oral contraceptive in that country. The products in Tables 1 and 2 are representative of products being sold throughout the world (Table 3).

Table 2. Low Dose Multiphasic Estrogen–Progestogen Combination Oral Contraceptives Marketed in the United States

--

Trade name	Ethinyl estradiol		Progestogen ^a		Launch date	Manufacturer
	Regimen, days	Dosage, µg	Regimen, days	Dosage, mg		
Triphasil	1–6	30	<i>Levonorgestrel</i>		12 84	Wyeth
	7–11	40	1–6	0.050		
	12–21	30	7–11	0.075		
Tri-levlen	1–6	30	12–21	0.125	1 86	Berlex
	7–11	40	1–6	0.050		
	12–21	30	7–11	0.075		
Jenest	1–21	35	12–21	0.125		Organo
			<i>Norethindrone</i>			
			1–7	0.50		
Ortho Novum 10 11	1–21	35	8–21	1.00	3 82	Ortho
			1–10	0.50		
Ortho Novum 7 7 7	1–21	35	11–21	1.00	4 84	Ortho
			1–7	0.50		
			8–16	0.75		
Tri-norinyl	1–21	35	17–21	1.00	4 84	Syntex
			1–7	0.50		
			8–16	1.00		
Ortho Tri-cyclen	1–21	35	17–21	0.50	10 92	Ortho
			<i>Norgestimate</i>			
			1–7	0.180		
			8–16	0.215		
			17–21	0.250		

^a Products are grouped by progestogen.

Table 3. Oral Contraceptives Marketed in Selected Countries

Trade name	Regimen, total days breakdown	Dose, mg and ingredients		Launch date	Manufacturer
		Progestogen ^a	Estrogen ^b		
Argentina					
Tridestan	21			9 82	Gador
	6	0.05 (11)	0.03 ^c		
	5	0.075 (11)	0.04 ^c		
	10	0.125 (11)	0.03 ^c		
Australia					
Biphasil	28			1 79	Wyeth
	11	0.05 (11)	0.05		
	10	0.125 (11)	0.05		
	7	placebo			
Neogynon	21	2.5 (11)	0.05	4 70	Schering
Austria					
Cilest ^d	21	0.25 (10)	0.035	8 89	Cilag
Gynovin	21	0.075 (9)	0.03	5 88	Schering
Micronovum ^e	28	0.35 (1)		9 72	Cilag
Ortho Novum 1 50 ^f	21	1.0 (1)	0.05 ^c	3 69	Cilag
Ovysmen 1 35 ^g	21	1.0 (1)	0.035	6 78	Cilag
Perikursal	21			6 77	Wyeth
	11	0.05 (11)	0.05		
	10	0.125 (11)	0.05		
Trinovum ^h	21			9 84	Cilag
	7	0.50 (1)	0.035		
	7	0.75 (1)	0.035		
	7	1.0 (1)	0.035		
Belgium					
Microgynon-50	21	0.125 (11)	0.05	9 67	Schering
Micronor ⁱ	28	0.35 (1)		3 73	Cilag
Microval	28	0.03 (11)		12 72	Wyeth
Brazil					
Anfertil	21	0.50 (11)	0.05	5 67	Wyeth
Evanor	21	0.25 (11)	0.05	8 70	Wyeth
Microdiol	21	0.15 (8)	0.03	7 85	Organon
Neovlar	21	0.25 (11)	0.05	2 71	Schering
Primovlar	21	0.50 (11)	0.03	6 67	Schering
Canada					
Micronor	28	0.35 (1)		5 72	Cilag
Nodestrin 1 50	21	1.0 (1)	0.05	8 64	Parke-Davis
Nodestrin 2.5 50	21	2.5 ^j	0.05	8 64	Parke-Davis
Ortho Novum 1 35	21	1.0 (1)	0.035	6 80	Cilag
Ortho Novum 1 50	21	1.0 (1)	0.035 ^c	6 62	Cilag
Ortho 7 7 7	21			9 83	Cilag
	7	0.50 (1)	0.035		

Ortho 10/11	7	0.75 (1)	0.035	11/81	Cilag
	7	1.0 (1)	0.035		
	21				
	10	0.50 (1)	0.035		
	11	1.0 (1)	0.035		
Adepal		France		10/76	Wyeth
	21				
	7	0.5 (11)	0.03		
Gynophase	14	0.20 (11)	0.04	1/74	Schering
	21				
	11	1.0 ^j	0.05		
Milligynon	10	2.0 ^j	0.05	7/78	Schering
	21	0.60 ^j			
	21	0.15 (11)	0.03		
Minidril	21			1/76	Wyeth
Miniphase	21			3/77	Schering
Ogyline	11	1.0 ^j	0.03	10/80	Roussel
	10	2.0 ^j	0.04		
	28	0.35 (1)			
	21	1.0 (1)	0.035		
	22				
Ortho Novum 1/35 ^k	7		0.05	9/81	Cilag
	15	2.5 (5)	0.05		
	21				
	6	0.05 (9)	0.03		
	5	0.07 (9)	0.04		
Ovanon	10	1.0 (9)	0.03	6/69	Organon
	22				
	7		0.05		
	15	2.5 (5)	0.05		
	21				
Phaeva	6	0.05 (9)	0.03	12/88	Wyeth
	5	0.07 (9)	0.04		
	10	1.0 (9)	0.03		
	22				
	7		0.05		
Physiostat	15	0.05 (5)		3/76	Organon
	21	2.0 (1)	0.05		
	21	1.0 ^j	0.03		
	21				
	7		0.05		
Planor	15	0.05 (5)		8/67	Roussel
	21	2.0 (1)	0.05		
	21	1.0 ^j	0.03		
	21				
	7	0.50 (1)	0.035		
Tretintovlane	7	0.75 (1)	0.035	11/75	Schering
	7	1.0 (1)	0.035		
	21				
	6	0.05 (9)	0.03		
	5	0.07 (9)	0.04		
Triella	10	0.10 (9)	0.03	7/84	Cilag
	21	0.15 (8)	0.03		
	21				
	7		0.05		
	15	0.05 (5)			
Triminulet	21	2.0 (1)	0.05	12/88	Wyeth
	6	0.05 (9)	0.03		
	5	0.07 (9)	0.04		
	10	0.10 (9)	0.03		
	21	0.15 (8)	0.03		
Vamoline		Germany		4/84	Organon
Anacyclin	22			5/70	Geigy
Conceplan M	16	1.0 (5)	0.05	9/78	Gruenenthal
	6	placebo			
	21	0.50 (1)	0.03		
	21	0.125 (11)	0.05		
	21	2.5 ^j	0.05		
Ediwal 21	21			1/63	Parke-Davis
Etalontin 21	21			11/66	Schering
Eugynon	21	0.50 (11)	0.05	10/75	Schering
Eunomin	21	2.0 ^j	0.1 ^c		Gruenenthal
Femranette	21	0.15 (11)	0.03	10/88	Brenner-Efeka
Lyndiol 2.5	22	2.5 (5)	0.05	1/62	Organon
Lyn-Ratiopharm	21	2.5 (5)	0.05	1/81	Ratiopharm
Marvelon ^m	21	0.15 (8)	0.03	6/81	Organon
Microlut	21	0.03 (11)		5/72	Schering
Neo-Eunomin	21	1.0 ^j	0.05	1/85	Gruenenthal
Noerdest 21	21	0.60 ^j	0.03	9/74	Parke-Davis
Noracyclin	22	2.5 (5)	0.05	1/64	Geigy
Orlest 28	28			7/67	Parke-Davis
Oviol	21	1.0 ^j	0.05	6/81	Nourypharma
	7	placebo			
	22				
	7	0.125 (8)			
	15		0.05		
Ovovesta	22	1.0 (5)		10/72	Organon
Ovysmen 0.5/35 ⁿ	21	0.50 (1)	0.035	3/75	Cilag
Pregnon 28	28			10/76	
Sequilar	22	1.0 (5)	0.05	3/74	Schering
	6	placebo			
	21				
	11	0.05 (11)	0.05		
	10	0.125 (11)	0.05		
Sinovula	21	1.0 ^j	0.05	3/73	Asche
Synfase	21			9/86	Gruenenthal
Tetragynon	7	0.50 (1)	0.035	9/85	Schering
	9	1.0 (1)	0.035		
	5	0.50 (1)	0.035		
	21	0.25 (11)	0.05		
	21				
Trinordiol	6	0.05 (11)	0.03	10/79	Wyeth

	5	0.075 (11)	0.04			
	10	0.125 (11)	0.03			
Triguilar	21			10	79	Schering
	6	0.05 (11)	0.03			
	5	0.075 (11)	0.04			
Tristep	10	0.125 (11)	0.03			
	21			10	83	Asche
	6	0.05 (11)	0.03			
	5	0.05 (11)	0.05			
	10	0.125 (11)	0.04			
		Italy				
Bivlar	21	0.50 (11)	0.05	4	80	Schering
Evanor D	21	0.25 (11)	0.05	2	73	Wyeth
Ginoden	21	0.075 (9)	0.03	10	87	Schering
Ovranet	21	0.15 (11)	0.03	9	78	Wyeth
Planum	21	0.15 (8)	0.03	4	84	Menarini
Practil 21	21	0.15 (8)	0.03	3	84	Organon
		The Netherlands				
Binordiol	21	0.05 (11)	0.05	2	75	Wyeth
Exluton	21	0.50 (5)		11	72	Organon
Modicon ^o	21	0.50 (1)	0.035	11	78	Cilag
Stediril	21	0.125 (11)	0.05	2	76	Wyeth
Trigynon	21			12	80	Schering
	6	0.05 (11)	0.03			
	5	0.075 (11)	0.04			
	10	0.125 (11)	0.03			
		Puerto Rico				
Brevicon	21	0.50 (1)	0.035	10	75	Syntax
Ortho Novum 7 7 7	21			6	84	Cilag
	7	0.50 (1)	0.035			
	7	0.75 (1)	0.035			
	7	1.0 (1)	0.035			
		South Africa				
Normovlar Ed	28			2	76	Schering
	11	0.05 (11)	0.05			
	10	0.125 (11)	0.05			
	7	placebo				
		Sweden				
Desolett	21	0.15 (8)	0.03	9	87	Organon
Follimin	21	0.15 (11)	0.03	3	76	Kabivitrurn
Follinett	21	0.25 (11)	0.05	9	71	Kabivitrurn
Follistrel	21	0.03 (11)		6	74	Kabivitrurn
Lyndiolett	21	1.0 (5)	0.05	6	77	Organon
Mini-Pe	21	0.35 (1)		3	72	Syntex
Neovletta	21	0.15 (11)	0.03	12	75	Schering
Regunon	21	0.125 (11)	0.05	9	76	Schering
Trionetta	21			6	81	Schering
	7	0.05 (11)	0.03			
	5	0.05 (11)	0.04			
	10	0.125 (11)	0.03			
		Switzerland				
Femovan	21	0.075 (9)	0.03	5	87	Schering
Gynera	21	0.075 (9)	0.03	4	87	Schering
Minulet	21	0.075 (9)	0.03	4	87	Wyeth
Ovostat	22	1.0 (5)	0.05	10	69	Organon
Stediril	21	0.50 (11)	0.05	1	66	Wyeth
Yermonil	21	2.0 (5)	0.04	8	73	Geigy
		United Kingdom				
Binovum	21			3	82	Cilag
	10	0.50 (1)	0.035			
	11	1.0 (1)	0.035			
Brevinor	21	0.50 (1)	0.035	1	77	Syntex
Conova 30	21	2.0 (4)	0.03	1	78	Gold Cross
Femodene	21	0.075 (9)	0.03	5	87	Schering
Gynovlar 21	21	3.0 ^l	0.05	10	64	Schering
Logynon Ed	28			6	80	Schering
	6	0.05 (11)	0.03			
	5	0.075 (11)	0.04			
	10	0.125 (11)	0.03			
	7	placebo				
Mercilon	21	0.15 (8)	0.03	4	88	Organon
Microgynon-30	21	0.15 (11)	0.03	3	74	Schering
Minovlar Ed	28			1	69	Schering
	21	1.0 ^l	0.05			
	7	placebo				
Minovlar 30	21	0.15 (11)	0.03	1	69	Schering
Norgeston	21	0.03 (11)		5	79	Schering
Noriday 28	21	0.35 (1)		10	72	Syntex
Norimin	21	1.0 (1)	0.035	11	78	Syntex
Ovran	21	0.25 (11)	0.05	11	72	Wyeth
Ovysmen	21	0.50 (1)	0.035	9	75	Cilag

^a Levonorgestrel (11); Norethisterone (1); Norgestimate (10); Lynestrenol (5); Ethynodiol diacetate (4); Desogestrel (8); Gestodene (9); Norgestrel (11).

^b Ethinyl estradiol unless noted.
^c Mestranol.
^d Belgium (8 89), France (9 88), Germany (11 86), Italy (10 90), Netherlands (12 90), Switzerland (1 87), UK (5 91).
^e Germany (11 71), Switzerland (7 74).
^f Belgium (9 69), Germany (6 70), Netherlands (6 71), Puerto Rico (1 63), Switzerland (7 67).
^g Germany (3 75).
^h Belgium (2 86), Germany (3 83), Italy (4 88), Mexico (7 84), Netherlands (2 85), Switzerland (3 84), UK (1 84).
ⁱ Puerto Rico (1 73), UK (10 72).
^j Norethisterone acetate [51-98-9], C₂₂H₂₈O₃.
^k Mexico (7 83), Puerto Rico (9 80).
^l Chlormadinone acetate [302-22-7], C₂₃H₂₉ClO₄.
^m UK (1 82).
ⁿ Switzerland (6 76).
^o Puerto Rico (1 75).

New, pharmacologically more selective progestogens, which attempt to eliminate undesirable pharmacological activity but retain the needed progestational activity, have been investigated. Studies published in the 1970s and 1980s demonstrate that androgenicity associated with the progestational component of some of the original progestogens is associated with changes in lipid metabolism (24,25). These changes may impact cardiovascular morbidity. Hence, oral contraceptives that do not disturb the various blood lipid fractions and do not lower HDL may be preferable to the ones that shift the lipid profile in an undesirable direction (26). Three companies, ie, Ortho Pharmaceutical Corporation, Organon, and Schering AG, have attempted to dissociate the androgenicity and progestational activities of steroidal progestogens using medicinal chemistry approaches. Norgestimate (10), desogestrel (8), and gestodene (9) emerged from this research (27–30). These more selective progestogens, in combination with ethinyl estradiol (12), compose the third generation oral contraceptives. During the 1990s and beyond, the usage of oral contraceptives containing these progestins will continue to grow.

Chemical Analysis. Chemically, the various progestogens belong to one of three classes. Estranes are 19-nortestosterone derivatives (Fig. 1); gonanes are 19-nortestosterone derivatives with a C-13 ethyl group (Fig. 2); and pregnanes are 17- α -OH progesterone derivatives similar in structure to progesterone itself.

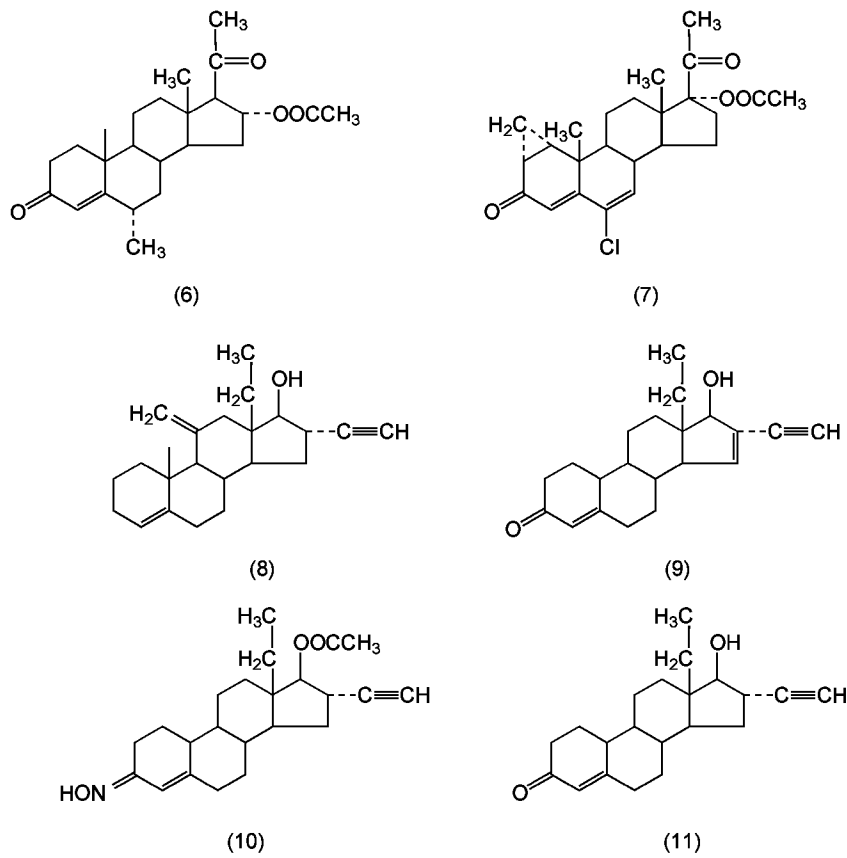
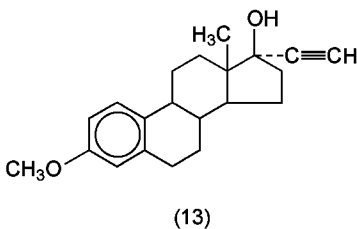
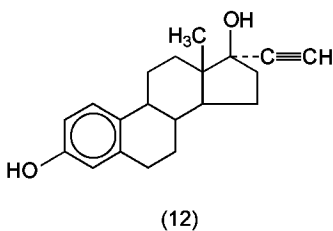


Fig. 2. Selected gonane progestogens. (8) desogestrel [54024-22-5], C₂₂H₃₀O; (9) gestodene [60282-87-3], C₂₁H₂ O₂; (10) norgestimate [35189-28-7], C₂₃H₃₁NO₃; (11) norgestrel [6533-00-2] (levonorgestrel, C₂₁H₂₈O₂).

The pregnanes include megestrol, medroxyprogesterone acetate [71-58-9], C₂₄H₃₄O₄ (6), chlormadinone acetate [302-22-7], C₂₃H₂₉ClO₄, and cyproterone acetate [427-51-0], C₂₄H₂₉ClO₄ (7).

The second active ingredient in combination oral contraceptives is estrogen. In the early 1990s, one of two related estrogens is utilized, ie, ethinyl estradiol [57-63-6] (19-nor-17-pregna-1,3,5(10)-trien-20-yne-3,17-diol), C₂₀H₂₄O₂, (12), and mestranol [72-33-3] (3-methoxy-19-nor-17-pregna-1,3,5(10)-trien-20-yne-17-ol), which has the molecular formula C₂₁H₂ O₂ (13).



Mestranol was the original estrogen contaminant discovered during contraceptive efficacy testing of the progestogen, norethynodrel. Drug metabolism studies indicate that most of the estrogenic activity of this compound could be attributed to its active metabolite, ethinyl estradiol, which has largely replaced mestranol as the estrogen component of oral contraceptives. Ethinyl estradiol differs from naturally occurring estradiol-17 β by having an ethinyl group attached to C-17, whereas mestranol differs from ethinyl estradiol by methylation of the hydroxyl group at C-3. Both ethinyl estradiol and mestranol resemble the natural estrogens qualitatively in their pharmacological actions on target tissues, including the reproductive tract, pituitary, and hypothalamus.

Progestogen-Only Oral Contraceptives. Progestogen-only oral contraceptives, ie, minipills, are available but are not used as extensively as combination oral contraceptives. These preparations contain 19-norsteroids such as norethindrone, lynestrenol, ethynodiol diacetate, or norgestrel, and 17- β -acetoxy progesterone derivatives such as chlormadinone acetate and megestrol acetate. The contraceptive effectiveness of these products is not as high as that of combination oral contraceptives; intermenstrual or breakthrough bleeding and spotting occur more frequently with these products. Progestogen-only oral contraceptives are prescribed for breast-feeding women since progestins do not influence milk production, and for women for whom estrogens are contraindicated.

Areas of Continued Research. Research continues in many academic and pharmaceutical laboratories throughout the world with the objective of improving oral contraceptives and better understanding their pharmacological and clinical actions.

Epidemiological studies with oral contraceptive users have had a bearing on the utilization of oral contraceptives by the public. The possible association of oral contraceptive use with the incidence of breast cancer has been the subject of numerous epidemiologic studies and reviews by the World Health Organization (WHO), National Institutes of Health (NIH), and the U.S. Food and Drug Administration (FDA). While the incidence of breast cancer in some small subgroups of users may have increased slightly, it has been concluded that the overall incidence in oral contraceptive users does not appear to have increased (31,32). Epidemiological studies on oral contraceptives have also demonstrated clear noncontraceptive health benefits, including protection against pelvic inflammatory disease (PID), ectopic pregnancy, endometrial cancer, cancer of the ovaries, benign breast disease, and benign ovarian cysts (33,34). Relief from a wide range of menstrual disorders, such as heavy menstrual flow, also is a beneficial effect of oral contraceptive use (29,30).

Toxicology studies are utilized to study the safety of new oral contraceptives. Because oral contraceptives are utilized by healthy, normal, young women, they have historically been required by regulatory agencies to undergo extremely intensive toxicological studies to confirm safety. Different countries have different requirements for the demonstration of product safety; the U.S. has historically been the most demanding, eg, safety had to be demonstrated in three animal species in very long-term chronic studies, utilizing different dosage schedules and different dosing intervals. For a number of years, issues such as species differences in pharmacokinetics and pharmacodynamics were not important considerations in the design of toxicological studies. A better understanding of the effect of kinetics, dynamics, and drug metabolism on drug safety has led to revisions in guidelines for the toxicological assessment of oral contraceptive safety (35,36). For example, use of the female beagle has traditionally been required for assessment of the carcinogenic potential of new progestins. A combination of laboratory and epidemiological studies has led to the conclusion that the female beagle cannot be used as a totally valid toxicological model for assessing the safety of new combination oral contraceptives (37). New FDA and WHO guidelines have been proposed.

Intensive research continues for new active ingredients that can be used in combination oral contraceptives. This research has been targeted at both new estrogens and more selective progestogens, including nonsteroidal agents; Norgestimate, Desogestrel, and Gestodene are all more selective pharmacologically than the original progestogens.

The potential use of natural estrogens, such as estradiol and estriol, as components of oral contraceptives has been described (38,39), but there has been little focus on the introduction of new natural estrogens into oral contraceptive formulations. The use of natural estrogens is based on the suggestion that the natural estrogens may have a lesser effect on coagulation factors and thus lesser predisposition to thrombus formation. The problem is that these natural steroids are poorly active as oral agents and require relatively high doses, in milligrams rather than micrograms, to provide adequate cycle control. Research continues in the synthesis of new, more selective steroidal estrogens that retain the endocrine activity necessary for control of ovulation and cycle control, but lack any other pharmacological activities. The potential exists to further dissociate these pharmacological actions through medicinal chemistry approaches.

A final area of research focuses on how the pill can best be used by women. Physicians used to recommend pill-free holidays, but it is now known that there is no valid reason for this practice. Similarly, further research indicated that the U.S. FDA's restrictive guidelines for prescribing the pill to women over the age of 35 are not justified, and those guidelines have been changed. Finally it is clear that cigarette smoking increases the risk of cardiovascular side effects, especially in women over 35 who utilize combination oral contraceptives.

Long-Acting Contraceptives

The establishment of fertility regulation with hormonally active oral contraceptives led to other routes of long-acting contraceptive drug administration. Long-acting contraceptives avoid compliance issues, and are useful in countries with fewer health professionals. Their popularity is based on simplicity of administration and a relatively high degree of effectiveness (see CONTROLLED-RELEASE TECHNOLOGY, PHARMACEUTICAL).

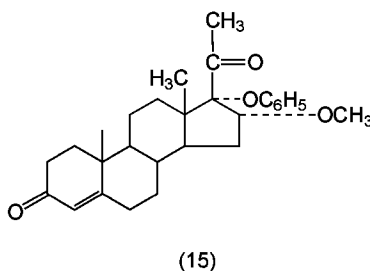
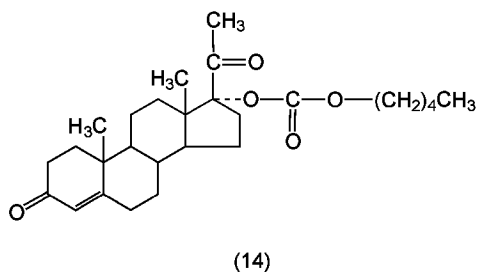
The currently (ca 1992) marketed injectable and implantable contraceptives are designed to be effective for maximum periods of three months and five years, respectively. There is little evidence from programmatic or health reasons that an injectable formulation with a longer effective life span, eg, six months, would not be equally effective. The acceptability and effectiveness of long-acting contraceptives may be determined by the means by which a community delivers contraceptive products to the public; the active life of a product may be determined by economic rather than programmatic or health related factors.

Injectable Contraceptives. Injections of contraceptive drugs must be sufficiently spaced to make the approach attractive to the user. The average contraceptive user is not willing to endure daily injections or injection intervals of one or two weeks; the minimum acceptable injection interval has been found to be at least four weeks.

Long-acting activity of injectable contraceptives can be accomplished via a number of different approaches. Steroid activity can be extended over a

period of time by chemical modifications, such as esterification, which result in products that require enzymatic hydrolysis over time for conversion to active products. Long-acting activity is also achieved by the presentation of drugs, with limited aqueous solubility, to the injection site in a microcrystalline aqueous suspension.

Two well-known injectable long-acting products are medroxyprogesterone acetate (Depo Provera) (6) and norethindrone enanthate (NET EN) (3). Both of these products are progestational in nature. Depo Provera (Upjohn Company) is a once-every-three-months injectable administered as a microcrystalline aqueous suspension. NET EN (Schering AG) is a two-month injectable administered as solution in castor oil. Depo Provera is approved for use as a contraceptive in more than ninety countries including the United States; NET EN is used in more than forty countries. Other products developed by chemically modifying known progestogens include 17 α -hydroxyprogesterone caproate [630-56-8], C₂₇H₄₀O₄ (14) and dihydroxyprogesterone acetophenide [24356-94-3], C₂₀H₃₀O₄ (15). These monthly injectables are widely used in Latin America.



Long-acting progestins act primarily as ovulation inhibitors. An important secondary component is their effect on the cervical mucus and endometrium, achieved at circulating blood drug levels below those required for ovulation inhibition (40).

The acceptability of long-acting injectable contraceptives is tempered by the effects of these drugs on endometrial function and the presence of nuisance side effects such as intermenstrual spotting and bleeding. Under normal circumstances the cyclical changes in endometrial morphology are the result of changes in ovarian function. Estrogens and progesterone, produced by the ovary, are responsible for the change from an estrogen dominated proliferative endometrium to the secretory progesterone dominated endometrium of the luteal phase. Diminution in progesterone production toward the end of a normal menstrual cycle leads to a breakdown of endometrial integrity and menses. Following the injection of long-acting progestins, ovarian function is inhibited and secretion of estrogens by developing follicles is diminished. Consequently, endometrial morphology no longer resembles that observed during menstrual cycles, and unpredictable bleeding episodes occur. This type of dysfunctional uterine bleeding frequently leads to the discontinuation of these contraceptive methods. Another outcome of the injectables is amenorrhea, ie, total absence of menstrual-like bleeding. Amenorrhea may be tolerated by patients better than dysfunctional bleeding, especially when this condition is explained to the patient prior to initiation of the therapy. However, amenorrhea is disturbing to some women because absence of menses is frequently associated with pregnancy.

The importance of predictable withdrawal bleeding has led to the development of combined progestin-estrogen injectable formulations. These products contain a relatively small dose of a long-acting progestin combined with a shorter acting estradiol ester. The progestational drug is programmed to have an effective life span of about thirty days and the estrogen of about a week. These types of formulations are administered on a monthly basis. Regularized bleeding occurs approximately one week after each injection and is the result of decreasing levels of estrogen. This is termed an estrogen withdrawal bleeding. Once-a-month injectables are popular in some Latin and South American countries; their introduction into other developing countries is being attempted by WHO (41).

The cost-effectiveness of injectable contraceptives has spurred the search for additional products that could be utilized for fertility regulation. The butanoate ester of levonorgestrel has undergone extensive studies in laboratory animals and has been the subject of clinical studies. The pharmacokinetic profile of this drug can be controlled, to a degree, by changing particle size distribution. A product containing an aqueous suspension of particles in the 15–20 μ m range can provide effective contraception for a three month period at a total dose of approximately 15 mg. This product is being developed by the World Health Organization (42).

Incorporation of a short-acting drug into a long-acting drug delivery system via microencapsulation technology is used to improve the pharmacokinetic profile of long-acting progestins. Biodegradable polymers such as polylactides and polyglycolides have been utilized to produce microcapsules or microspheres with the polymer surrounding the drug (42). The kinetic profile of such preparations approaches zero order release for the desired period of time. Most of these formulations have been programmed to release the progestin for approximately three months. While clinical studies with norethindrone-releasing microcapsules have been ongoing for a number of years, reproducibility of the product has sometimes been a problem. Improved manufacturing processes should eliminate batch to batch variations. Finally, the cost-effectiveness of delivery via microencapsulation of a product with superior pharmacokinetic characteristics is yet to be documented. The overall process of delivering a microencapsulated contraceptive to the client is more complex than for a micronized aqueous suspension. Consequently the improvement in pharmacokinetic profile may have to be weighed against higher cost and dosage form complexity.

Implanted Contraceptives. Controlled release of contraceptive progestins also can be accomplished by incorporating the drug into an implantable cylinder or rod. The best known implant is Norplant, developed by the Population Council. Norplant is composed of six matchlike silastic cylinders with the progestogen levonorgestrel (11) incorporated on the inside of each cylinder. After implantation under the skin, the product can provide effective contraception for a period of five years (43,44). Since silastic is not biodegradable, the implant can be removed from a patient at any time.

Clinical studies with Norplant attest to its high contraceptive efficacy and safety. The main reason patients request the removal of Norplant is unpredictable vaginal bleeding episodes followed by amenorrhea. The bleeding problem is an unavoidable sequela of progestogen-only contraception.

While the effective life of Norplant I is at least five years, most of the users have the implants removed at an earlier time. Norplant II was developed to have a shorter life-span, and is composed of only two rods. Levonorgestrel is dispersed in the silastic matrix which is then inserted into a thin silastic tubing. This product has an effective life-span of two to three years (36). Because it is composed of only two rods, Norplant II also is easier to implant and remove.

Clinical studies also have been carried out with nonbiodegradable implants releasing the progestin desogestrel (8). Unlike levonorgestrel, desogestrel possesses lower androgenic activity and thus has less adverse effect on blood lipids.

Biodegradable Implants. Utilization of biodegradable polymers obviates the need for implant removal. However, biodegradation should not take place before the drug release is essentially finished; before that, structural integrity permitting surgical removal of the implant must be maintained. The time-course for biodegradation and the disappearance of the implant are still being studied.

The Capronor device has walls composed of sigma caprolactone and releases levonorgestrel. It is a single implant with projected life span of 12–18 months (42). Capronor I had the drug dissolved in ethyl oleate. Once the ethyl oleate diffused through the device wall, the rate of drug release decreased to levels below those required for ovulation inhibition. Clinical studies carried out by the National Institute for Health (NIH), WHO, and the Indian Council of Medical Research indicated that the presence of ethyl oleate reduced the useful life of the device to 8–10 months. Since the use of excipients such as ethyl oleate is undesirable, investigators have attempted to regulate the rate of drug release by reducing the wall thickness. Studies in animals indicate that the thinner walled Capronor II may have an effective life of at least 12 months.

Fused Pellets. Another form of an implant is a fused pellet. The pellets may be composed of either the drug alone, or the drug fused with cholesterol (29,36), and are formed as small cylinders by melting the drug and then solidifying it under pressure. Clinical studies with norethindrone pellets have been in progress for a number of years. Effective rates of release of the drug from the implantation site were originally difficult to achieve.

Polymeric implants and pellets require minor surgery for both their insertion and removal.

Vaginal Rings. Vaginal epithelium is readily permeable to contraceptive steroids. Since the vascular drainage of the vagina bypasses the liver, this route of administration potentially permits utilization of drugs that have low oral activity.

Contraceptive vaginal rings consist of silastic shells or core rings of various sizes and membrane thicknesses. They have been developed for delivery of progestins alone or progestins combined with estrogens. Progestin-alone rings are conceptionally similar to implants, but are under direct control of the patient since they can be removed at any time. Rings releasing both a progestin and an estrogen resemble combined oral contraceptives. After the ring is removed, a predictable withdrawal bleeding takes place. Contraceptive efficacy of vaginal rings has been linked to the weight of the subject. Higher release rates are required for heavier women than for lower weight women. Vaginal rings have been under development for nearly fifteen years, by the World Health Organization and Population Council (45,46). Acceptability of this route of contraceptive drug administration in developed countries must still be determined.

Contraceptational Drugs

Pharmacological substances that either inhibit implantation or interrupt pregnancy after implantation have been investigated during the 1980s and 1990s. A number of different terms have been used to describe these compounds (47–63), including anti-implantive agents, postcoital contraceptives, morning-after pills, once-a-week pills, interceptives, abortifacients, and contragestational agents. This medical approach to fertility regulation presents several principal advantages, including potentially fewer long-term side effects as a result of short-term periodic administration, and greater convenience.

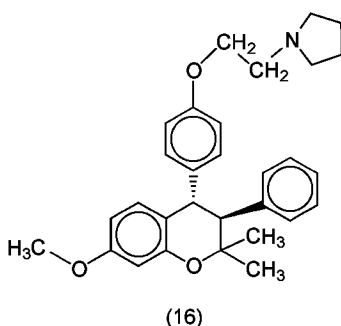
However, in the decision to develop and market contragestational drugs, social, political, legal, ethical, and religious factors have been of critical importance.

Post-Coital Contraception. Post-coital contraception historically has been viewed as an emergency measure where regular contraceptives were not used or where the primary contraceptives may have failed. A number of different approaches have been utilized in post-coital regimens. Early regimens utilized high doses of the estrogens (63), diethyl stilbestrol [56-53-1], ethinyl estradiol (12), and the conjugated equine estrogens (Premarin) (7). Characteristically, the drugs are taken for a period of three days. When given later than 72 hours after coitus, the effectiveness is reduced; the drugs are ineffective if implantation has been established. Although the estrogens are highly effective, users suffer from a high incidence of nausea and vomiting; cycle regularity also is disturbed. A similar approach has been reported for levonorgestrel-containing contraceptives (64).

It has been demonstrated that Danazol [17230-88-5], $C_{22}H_{27}NO_2$, is highly effective in post-coital regimens, and has a very low incidence of side effects. Danazol is marketed in many countries for the treatment of endometriosis and its availability is unrestricted. It does not appear that there is a significant difference in effectiveness when doses of 800 to 1200 mg have been utilized daily for three days (64,65). When utilized in the post-coital mode, the precise mechanism of action of Danazol is not well-understood.

Clinical studies with the antiprogestin RU-486 indicate that, when used in a post-coital mode, this drug may be effective in preventing pregnancy.

A nonsteroidal weekly pill, centchroman [31477-60-8] (16), has recently been launched by the Central Drug Research Institute, Lucknow, India, and is being marketed as Choice-7 and Sahali (Hindustan Latex). Centchroman, (3,4-*trans*-2,2-dimethyl-3-phenyl-4-[*p*-(β -pyrrolidino-ethoxy)-phenyl]-7-methoxy-chromane) inhibits implantation of the fertilized egg, thus avoiding pregnancy (66). It exerts its antifertility effect via weak estrogenic and potent anti-estrogenic activity (67). Although the synthesis and pharmacological actions of centchroman have been well-documented (68), overall efficacy and side effects are not known or described except in a brief description of the product. It is reported that the Pearl index has been calculated to be 3.05 when weekly doses of centchroman (30 mg) were administered to approximately 1,600 women for a total of 20,000 months (66).



Abortifacients. It is estimated that between 30 and 40 million legal abortions and the same number of illegal abortions are carried out each year worldwide (69). In the mammal, removal of the corpus luteum during early pregnancy results in termination of pregnancy. Pharmacologically, this can be accomplished by various methods, including inhibition of gonadotropin support of the corpus luteum, inhibition of progesterone biosynthesis in the ovary, inhibition of progesterone binding to the uterine progestin receptor, and increase in the metabolism of progesterone. Many compounds that affect these processes have been reported (71); some of these compounds have been introduced for clinical use. Chemically induced abortion originally involved the administration of the natural prostaglandin $F_2\alpha$ [551-11-1] (Dinoprost, $C_{20}H_{34}O_5$) (17), or synthetic analogues such as sulprostone [60325-46-4], $C_{23}H_{31}NO_7S$ (18) and prostaglandin ONO 802, [64318-79-2] (Gemeprost, $C_{23}H_{38}O_5$) (19) (Fig. 3) (71). After administration of relatively small amounts of these prostaglandins, the muscular tone of the uterus increases, followed by contractions, cervical dilation, and expulsion of the uterine contents. The usefulness of prostaglandins for termination of pregnancy is limited, because of a high frequency of gastrointestinal side effects.

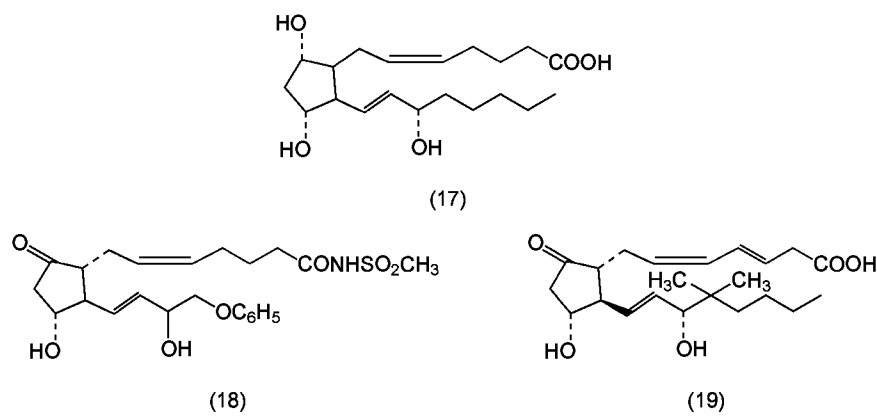


Fig. 3. Prostaglandins used in chemically induced abortion.

In 1988, the French government approved the marketing of an abortion pill. RU486 [84371-65-3] (Mifepristone) (20), an antiprogesterone developed by the French pharmaceutical company Roussel-UCLAF, is the first clinically useful progesterone antagonist (Fig. 4). A review of the chemistry, pharmacology, and clinical applications of this compound (72) is available, as is a review on the use of RU486 alone or in combination with a prostoglandin analogue for termination of early pregnancy (73). Other studies suggest the use of antiprogesterins for contraception and for treatment of gynecological disorders related to hormone production (74). The discovery of RU486 was followed by laboratory and clinical studies with other antiprogesterins, including ZK98,299 (22) and ZK98,734 (21) (Fig. 4) (76). Many issues, not only scientific, but also political and religious, surround the clinical application of progesterone antagonists (5), and it is difficult to project worldwide availability of RU486 and related products.

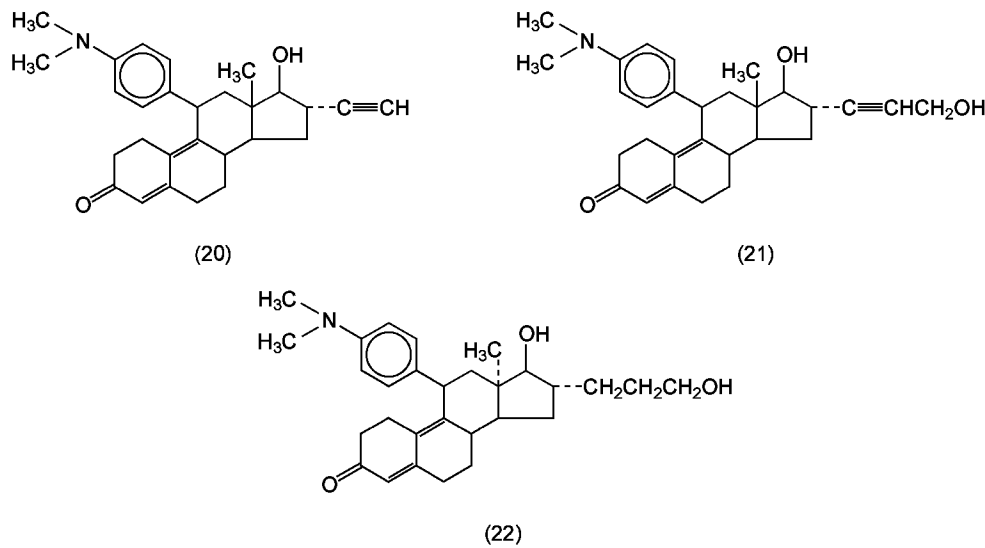


Fig. 4. Progesterone antagonists, ie, antiprogesterins. (20) RU 486; (21) lilopristone [97747-88-1] (ZK 98, 734); (22) onapristone [96346-61-1] (ZK 98, 299).

Another approach to pregnancy termination has been the utilization of progesterone synthesis inhibitors. These compounds block production of progesterone by the corpus luteum and the early placenta. When administered with a prostaglandin, their use can result in medically induced abortions.

Vaginal Contraceptives

Vaginal contraception dates back to 1850 B.C. when written instructions for vaginal contraceptives appeared in Egyptian papyri. Vaginal contraceptives are simple, safe, require little medical supervision, and are among the most widely used forms of birth control available. Where used consistently and properly, reasonable efficacy rates are achieved (76,77). The disadvantages of this method are that its use is coitally related and its efficacy depends on proper usage.

In 1855, the effects of chemical agents on sperm motility (78) were reported, prompting the first studies in modern-day vaginal contraceptive research. In 1880, the first commercial vaginal contraceptive, a suppository of cocoa butter containing the spermicide quinine sulfate [804-63-7] was produced (79). Subsequently, several vaginal gels and suppositories containing a variety of spermicidal chemicals, for use alone or in combination with a diaphragm, were marketed throughout the world. The International Planned Parenthood Contraceptive Directory for 1981 lists over 100 vaginal contraceptive products available worldwide (80).

The active agents employed in vaginal contraceptive products may be classified as weak acids, organometallic compounds, and surfactants (81). Because of the inferior spermicidal potency of the weak acids, and the growing concerns over the potential toxicity of mercury-containing compounds, surface active agents constitute the most important class of spermicidal compounds in vaginal contraceptive products. A review of the history of vaginal contraceptives that provides description of selected products and reviews the literature on clinical efficacy and various aspects of use and distribution is available (1).

The basis for *in vitro* spermicidal assays, used to evaluate new spermicidal agents, was introduced in 1932 (82). The Sander Cramer test (83) (Ortho Pharmaceutical Corporation), developed in the late 1930s, has been widely used to screen new spermicidal agents and compare spermicidal formulations. In the 1940s, the first surface active spermicidal agents, nonoxynol-9 [26027-38-3] (7-nonylphenoxy polyethoxy ethanol, Triton N) (23) and octoxynol [9002-93-1] (4-diisobutylphenoxy polyethoxy ethanol, Triton X-100) (24) were developed at Ortho (84). During the next two decades, these two agents became the principal active ingredients utilized in vaginal contraceptives throughout the world. A third surfactant spermicidal agent, menfegol [57821-32-6] (4-menthanylphenyl polyoxyethylene [8,8] ether) (25) was discovered and developed in the late 1960s (Fig. 5) (85).

hysterectomy	6	5	6	5
menopause				
withdrawal	4	5	5	5
rhythm	3	3	4	3
diaphragm	4	4	3	3
pregnant	3	3	3	3
trying to conceive	2	2	3	3
sponge	2	2	2	2
vaginal suppository	2	2	2	2
foam	1	1	2	1
IUD	1	1	1	1
douche	1	1	1	1
cream		1	1	1
jelly alone				
cervical cap				
Norplant	na	na	na	na
no method	24	22	20	19

^a *Courtesy of Ortho Pharmaceutical, Raritan, N.J.*

In the 1970s questions were raised about certain immunological complications as a consequence of male sterilization or vasectomy. Clinical epidemiological data do not appear to indicate that this actually occurs in clinical practice. No significant long-term side effects of male sterilization have been demonstrated.

Voluntary female sterilization is the world's most widely used family planning method. An estimated 138 million women of reproductive age used the method in 1990, 43 million more than in 1984. Millions more are expected to ask for the method during the 1990s (97).

Because of the increasing worldwide interest and demand for simple, effective, and inexpensive female sterilization, a variety of procedures and methods have been developed. These approaches differ whether they are performed postpartum, postabortum, or in interval situations. The choice of methods also largely depends upon the physician's prior training, knowledge, and experience. Excellent reviews have been written on sterilization (98).

Physical Barrier Methods

Various physical barrier devices are available for contraceptive use by men and women. Modern barrier methods such as diaphragms, condoms, and cervical caps were made possible by the discovery of the vulcanization of rubber.

Diaphragms, shallow rubber cups with a flexible metal rim which are placed in the vagina and cover the cervix, are both a mechanical barrier to sperm and a receptacle for spermicidal agent. The mechanism of action is believed to be a combination of spermicidal and mechanical barrier actions. To be effective, the diaphragm must fit correctly, be inserted properly, and remain in place sufficiently long for the spermicide to act (99).

The cervical cap birth control device has been available in Europe for many years and in the U.S. since late 1988. It is a small, rubber, dome-shaped device that fits snugly over the cervix. The cervical cap has some advantages over the diaphragm, but has not lived up to widespread expectations that it would become an overwhelmingly popular method of contraception (100).

The male barrier contraceptive device is known as the condom, or rubber, and is widely available in most countries. The condom is a rubber or latex sheath, sometimes packaged with a lubricant and spermicide, which serves as a cover for the penis and a receptacle for semen. The method is very effective if the condom is of good quality, remains on, and is replaced for each subsequent intercourse. It was reported that 6 billion condoms were used in 1990 (101). Usage appears to be increasing as adjunctive use with other methods of contraception for prevention of HIV or other sexually transmitted diseases. By rough estimate, condoms may have been used in more than 13 billion acts of sexual intercourse that risked unwanted pregnancy, HIV, and or other sexually transmitted diseases (101).

Two new female condoms, ie, vaginal pouches, are in early stages of development. These devices still require thorough preclinical and clinical studies to demonstrate safety and effectiveness before they reach the market (102).

Another type of barrier contraceptive device is the vaginal contraceptive sponge. Clinicians do not appear to be hailing this method. Studies show that the failure rate for the contraceptive sponge in parous women is higher than the failure rate for the diaphragm. A U.S. study of the contraceptive sponge found a first-year failure rate of 13.9% in nulliparous sponge users and a 28.3% first-year failure rate in parous sponge users. Other disadvantages of the method include allergic reactions in a small percentage of women, and a slightly increased risk of nonmenstrual toxic shock syndrome (TSS). When compared with other over-the-counter methods, the sponge can be expensive (103).

Natural Methods

Natural methods, ie, natural family planning, are methods based on awareness of the fertile and infertile segments of the menstrual cycle. This awareness can be utilized to avoid pregnancy or to become pregnant.

In order to avoid conception, abstinence from sexual intercourse during the fertile period of the menstrual cycle must be practiced (77,104). It has been determined that the fertile period in women occurs before menstruation (105,106), and formulas have been developed to determine the fertile and infertile days of the menstrual cycle. Ovulation has been linked to a cyclic shift in basal body temperature (107), which can be used retrospectively to determine the time of ovulation.

The primary difficulty with periodic abstinence is the month-to-month variation in the time of ovulation. Whereas the ovum can only be fertilized during the first 12 to 24 hours after its release from the ovary, sperm remain viable longer in the female reproductive tract, able to fertilize an ovum for 5–7 days and perhaps longer. Thus, intercourse several days prior to ovulation can result in pregnancy.

Recent findings may lead to better identification of the fertile period. Changes in the quality and quantity of cervical mucus occurring 5–6 days prior to the mid-cycle surge of luteinizing hormone (LH), which initiates ovulation, can be used to predict the fertile period. However, it is sometimes difficult to recognize the changes in mucus. Various devices have been marketed to automate the daily temperature monitoring or to assist in cervical mucus collection (68). However, such devices produce only marginal improvement because of the inherently variable nature of temperature and mucus texture as fertility markers. Colorimetric enzyme immunoassays have been developed for the measurement of LH in urine. Since LH is rapidly excreted, the increase in urine LH levels can be used as a marker to predict impending ovulation. However, because of the life span of sperm, one must be cautious in predicting the usefulness of this method. Research on methods of consistently and accurately predicting ovulation is ongoing (108).

Breast-Feeding

In many societies, it is believed that women who are breast-feeding are incapable of becoming pregnant. Suckling leads to a release of prolactin and endorphins that interfere with the hormones necessary for ovulation. This disruption of ovulation lasts for several postpartum months; in some populations it may last as long as a year or more. A study of lactating women in the United States and the Philippines concluded that during the first six months postpartum women who were fully breast-feeding had only between 1 and 5% risk of ovulation; women who were partially breast-feeding had less than a 10% risk (109). However, although breast-feeding does affect ovulation, the duration of lactational amenorrhea and infertility is variable (110), and lactation appears to be unreliable as the sole method of fertility regulation.

New Approaches

There continues to be great interest in developing new and improved contraceptives. In addition to utilizing new technology, new contraceptives should be superior to existing products, eg, oral contraceptives, used by millions of women over the last 30 years, are not only safe and effective but even protect women against some cancers. Because oral contraceptives are so effective, they become a very high standard that other products must meet. However, improved methods of contraception are still needed by segments of the world's population.

Contraceptive Vaccines. Major research efforts involve immunological approaches to fertility control; excellent reviews of this area are available (111–114). The development of contraceptive vaccines is directed towards the immunoneutralization of reproductive process or the interference of fertilization by inducing antibodies against oocytes and spermatozoa. Attempts have been made to develop vaccines against leutinizing hormone releasing hormone (LHRH) (also known as gonadotropin releasing hormone, GnRH), LH, follicle stimulating hormone (FSH), human chorionic gonadotropin (hCG), placenta antigen, the zona pellucida of the ovum, and different sperm antigens.

Research on an hCG vaccine has been conducted over the past 15 years. WHO has conducted a phase I clinical study in Australia, using a vaccine based on a synthetic C-terminal peptide (109–141) of β -hCG conjugated to Diphtheria Toxoid (CTP-DT), that showed potentially effective contraceptive levels of antibodies were produced in vaccinated women without any adverse side effects. Phase II clinical studies are under consideration to determine if the immune response, raised to its prototype anti-hCG vaccine, is capable of preventing pregnancy in fertile women volunteers (115). While research on the C-terminal peptide from the β -subunit of hCG has been carried out under the auspices of WHO, research supported by the Population Council and the National Institutes of Health has involved two alternative vaccine candidates (109,116,118).

Using recent advanced technologies, unique sperm antigens have been identified and partially characterized. Sperm antigens shown to have high immunocontraceptive potential are human sperm membrane antigen (SP-10) and guinea pig sperm membrane protein (PH-20). SP-10 is a sperm membrane-specific antigen of 24-34 kD, isolated using a monoclonal antibody (MHS-10) that cross-reacts with the entire acrosomal region. It is associated with the outer aspect of the inner acrosomal membrane and the inner aspect of the outer acrosomal membrane of mature human sperm (119). It has been produced recombinantly in an *Escherichia coli* expression system. The recombinant SP-10 fusion protein is under study in the baboon.

PH-20, a guinea pig sperm protein of 64 kD, is present on both the plasma membrane and inner acrosomal membrane of sperm. It is essential for adhesion of sperm to the zona pellucida, the initial step in the fertilization process. Active immunization with PH-20 causes infertility in both male and female guinea pigs for a period ranging from 6 to 15 months (120).

Another interesting sperm specific antigen is lactic dehydrogenase-x (LDH-x or LDHC_x), an isoenzyme of LDH confined to male germ cells. LDH-x is one of the best characterized antigens and its amino acid sequence is known. A synthetic peptide based on a portion of the molecule has been shown to reduce fertility in laboratory animals. The nucleotide sequence coding for human LDH-x has been defined and engineered into an expression vector system (121).

The zona pellucida (ZP) is the complex extracellular glycoprotein matrix that surrounds the oocyte. It plays an important role in sperm penetration and fertilization. It is a composite of several antigenic glycoproteins designated ZPI, ZPII, ZPIII, and, in some species, ZPIV. Immunologically, it does not cross-react with any other body tissues, but interspecies cross-reactivity has been observed among several species including primates. Numerous studies indicate the immunocontraceptive potential of zona pellucida (122).

Several other antigens with good immunocontraceptive potential have been identified and investigated in laboratory animals. In most studies, the rate and duration of the immunocontraceptive effect are less than acceptable. A potential problem in immunological approaches to antifertility research is the need for a safe, effective adjuvant and suitable animal models for evaluating the efficacy and safety of methods (111). Newer and more effective adjuvants are required for contraceptive vaccines and vaccines in general.

Luteinizing Hormone Releasing Hormone. The isolation and synthesis of luteinizing hormone releasing hormone (LHRH) was an important advance in reproductive research (123). LHRH is a peptide hormone produced and secreted by the hypothalamus that stimulates the secretion of FSH and LH. A decapeptide with an amino acid sequence of pyroGlu-His-Trp-Ser-Tyr-Gly-Leu-Arg-Pro-Gly-NH₂, it is chemically and functionally similar in both males and females of all mammalian species studied thus far. More than 1000 analogues of LHRH have been synthesized by deletion or substitution of amino acids to introduce agonist or antagonist properties. A large number of reviews have appeared discussing their therapeutic importance (124–126).

These agonists and antagonists may provide useful therapy for several clinical conditions such as prostatic carcinoma, precocious puberty, and endometriosis. Scientists also continue to study LHRH analogues for contraception. Treatment with LHRH analogues blocks ovulation in women and spermatogenesis in men. However, it also results in a loss of estrogen and testosterone and causes other related side effects. Scientists are attempting to assess the use of these agents in combination with replacement estrogen, progesterone, and testosterone to determine if the side effects can be avoided. Several small nonhuman primate and clinical studies have suggested possible utility in this area, but large-scale clinical efficacy studies have not been conducted.

Inhibin and Activin. Inhibin, a water-soluble, gonadal factor known for over 50 years to inhibit pituitary function, has been isolated and identified (127–130). Inhibin is a glycoprotein hormone that preferentially inhibits the secretion of FSH. It consists of an α -chain subunit, mol wt 14,000, linked by disulfide bonds to a β -chain subunit, mol wt 18,000. There exist two forms of the β -chain subunit, β -A and β -B. The smaller subunit combines with either the β -A or β -B subunit to form inhibin-A or inhibin-B, respectively.

During the isolation of inhibin from follicular fluid, some chromatographic fractions stimulated FSH release from cultured anterior pituitary cells, suggesting the existence of FSH releasing proteins (FRPs). Two FRPs, given the generic term activins, were subsequently isolated (131,132). One is composed of two disulfide-linked β -A subunits (activin A); the other consists of similarly linked β -A and β -B subunits (activin AB).

Studies confirm that inhibin plays a role in regulating FSH secretion. However, the importance of this role in the human has not yet been determined. If inhibin-regulated FSH secretion is pivotal in follicular recruitment and growth, then it may be possible to block ovulation by means of inhibin antagonists.

Activin has the potential to serve the same therapeutic uses as GnRh analogues. This is because of its ability to suppress steroidogenesis directly at the gonadal level. Although the spectrum of functions of inhibin and activin are not completely understood at present, this peptide family has already demonstrated, by the nature of its differential subunit association, a powerful mechanism for the generation of dimers with opposing biologic actions. These characteristics of the inhibin peptide family warrant further study and evaluation as alternative approaches to fertility control.

Progesterone Antagonists as Contraceptives. Another area of antifertility research involves progesterone antagonists or inhibitors. This hormone is required to maintain pregnancy, and infertility results from failure of the corpus luteum to produce adequate amounts of progesterone. Inhibitors of progesterone synthesis, such as epostane (133), and inhibitors of progesterone-receptor binding, such as RU486, have been investigated for termination of pregnancy. Studies in the nonhuman primate indicate that progesterone antagonist may have antifertility potential other than as an abortifacient (65).

Male Fertility Control. There is interest in male fertility control, both from a scientific as well as a sociological viewpoint. Many compounds have been identified as having male antifertility activity in various species, eg, gossypol [303-45-7], ORF 5513, 5-thio-D-glucose [20408-97-3], and 6-chlorodeoxyglucose (134). A principal program centering around the use of androgens has been conducted (135).

Organic molecules thus far identified, such as those listed above, appear either to have irreversible antifertility effects, to be inherently toxic, or to affect libido. It has been demonstrated that sperm count could be depressed in men injected with large doses of androgens. However, questions about the potential utility of androgens as male antifertility agents are still debated.

The ideal male contraceptive would produce azoospermia without compromising libido and sexual potency. While not totally fulfilling the criteria for a perfect male contraceptive, GnRH antagonists hold a greater potential than GnRH agonists. Unlike the agonists, GnRH antagonists inhibit gonadotropin secretion, decrease androgen levels, and induce azoospermia in male primates (136,137). Similar effects on hormone secretion have been reported in men (130).

In monkeys, testosterone replacement delays, but does not prevent, GnRH antagonist-induced azoospermia (139). In men, the combination of testosterone and a GnRH antagonist results in a more complete gonadotropin and gonadal suppression than either agent alone (140). These results suggest

that GnRH antagonists, given in conjunction with androgens to maintain libido and sexual potency, have potential as male contraceptives.

BIBLIOGRAPHY

"Contraceptive Drugs" in *ECT* 2nd ed., Vol. 6, pp. 60–92, by F. B. Colton and P. D. Klimstra, G. D. Searle & Co.; in *ECT* 3rd ed., Vol. 6, pp. 743–783, by L. J. Chinn and F. B. Colton, G. D. Searle & Co.

1. R. Belsky, *Vaginal Contraceptives: A Time for Reappraisal? Population Reports*, series H, No. 3, The George Washington University Medical Center, 1975.
2. C-R. Garcia and D. L. Rosenfeld, *Human Fertility: The Regulation of Reproduction*, FA Davis Company, Philadelphia, 1977, p. 59.
3. L. Mastroiani, Jr., P. J. Donaldson, and T. T. Kane, *Developing New Contraceptives*, National Academic Press, Washington, D.C., 1990, p. 11.
4. R. O. Greep, M. A. Koblinsky, and F. S. Jaffe, *Reproduction and Human Welfare*, MIT Press, Cambridge, Mass., 1976, p. 53.
5. B. Hedon, *Acta Obstet. Gynecol. Scand.* **69**, Suppl. 152, 7 (1990).
6. E. Diczfalusy, *Int. J. Fert. Suppl.*, 56 (1989).
7. R. A. Hatcher, F. Stewart, J. Trussel, and co-workers, *Contraceptive Technology*, Irvington Publishers, Inc., New York, 1990–1992, p. 227.
8. G. Pincus, *The Control of Fertility*, Academic Press, New York, 1965.
9. C. Djerassi, in M. J. Parnham and J. Bruinvels, eds, *Discoveries in Pharmacology*, Vol. 2, Elsevier, Amsterdam, 1984, p. 339.
10. L. Haberlandt, *Munch. Med. Wochenschr.* **68**, 1577 (1921).
11. C. R. Kay, *J. Royal College of Gen. Practitioners* **30**, 8 (1980).
12. W. H. W. Inman, M. P. Vessey, B. Westerholm, and A. England, *Br. Med. J.* **2**, 203 (1970).
13. Royal College of General Practitioners, *Oral Contraceptives and Health: An Interim Report from the Oral Contraception Study*, Pitman Medical, New York, 1974.
14. M. P. Vessey, R. Doll, R. Peto, and co-workers, *J. Biosoc. Sci.* **8**, 373 (1976).
15. A. Rosenfield, *Am. J. Obstet. Gynecol.* **132**, 92 (1978).
16. H. W. Ory, A. Rosenfield, and L. C. Landman, *Fam. Plann. Perspect.* **12**, 278 (1980).
17. L. E. Bottiger, G. Boman, G. Elkund, and B. Westerholm, *Lancet* **I**, 1097–1101 (1980).
18. S. Ramcharan, R. A. Pellegrin, R. M. Ray, and co-workers, *The Walnut Creek Contraceptive Drug Study: A Prospective Study of the Side Effects of Oral Contraceptives*, Vol. 3, U.S. Dept. of Health and Human Services, Government Printing Office, 1981.
19. M. Notelovitz, C. S. Kitchens, L. Coone, L. McKenzie, and R. Carter, *Am. J. Obstet. Gynecol.* **141**, 71–75 (1981).
20. T. W. Meade, *Am. J. Obstet. Gynecol.* **142**, 758–761 (1982).
21. D. R. Mishell, Jr., *Am. J. Obstet. Gynecol.* **142**, 809 (1982).
22. "Oral Contraceptives", in *Popul. Rep.* [A] **6** (May–June 1982).
23. S. A. Pasquale, *J. Reprod. Med.* **29** (Suppl.), 560 (1984).
24. K. Fotherby, *Br. J. Family Planning* **11**, 86 (1985).
25. G. Silfverstolpe, A. Gustafson, G. Samsoie, and A. Svanborg, *Acta Obstet. Gynecol. Scand.* **88** (Suppl), 89 (1979).
26. T. Gordon, W. P. Castelli, M. C. Hjortland, and co-workers, *Am. J. Med.* **62**, 707 (1977).
27. A. Phillips, D. W. Hahn, S. Klimek, and J. L. McGuire, *Contraception*, **36**, 181 (1987).
28. A. Phillips, K. Demarest, D. W. Hahn, and co-workers, *Contraception* **41**, 399 (1990).
29. H. J. Kloosterboer, C. A. Von Koordegraaf, and E. W. Turpijn, *Contraception* **38**, 325 (1988).
30. D. W. Hahn, A. Phillips, and J. L. McGuire, *Aktuelle Aspekte der Hormonalen Kontrazeption*, Karger, New York, 1991, p. 46.
31. World Health Organization, *Oral Contraceptives and Neoplasia*, WHO Technical Report Series, 817, Geneva, 1992.
32. R. Derman, *J. Reprod. Med.* **31** (9), Suppl. 879 (1986).
33. *Oral Contraceptives in the 1980s*, *Population Reports*, Series A no. 6, Population Information Program, The Johns Hopkins University, Baltimore, 1982.
34. A. Rosenfield, *Contemporary OB/GYN*, 136, (1983).
35. A. T. Gregoire and R. P. Blye, eds., *Contraceptive Steroids: Pharmacology and Safety*, Plenum Press, New York, 1986.
36. F. Michal, ed., *Safety Requirements for Contraceptive Steroids*, Cambridge University Press, Cambridge, 1980.
37. *World Health Organization Collaboratory Study of Neoplasia and Steroid Contraceptives*, *Lancet* **338**, 833 (1991).
38. G. Samsioe, A. Skryten, and G. Silverstolpe, *Gynecol. Obstet. Invest.* **15**, 275 (1983).
39. E. Bostofte, L. Hemmingsen, K. Y. Alling-Moller, and co-workers, *Acta Endocr. Copenh.* **87**, 855 (1978).
40. *Hormonal Contraception: New Long-Acting Methods*, *Population Reports*, Series K no. 3, Population Information Program, The Johns Hopkins University, Baltimore, 1987.
41. World Health Organization *Annual Report*, 1991.
42. G. L. Zatuchni, A. Goldsmith, J. D. Shelton, and J. J. Sciarra eds., *Long-Acting Contraceptive Delivery Systems*, Harper and Row, Philadelphia, 1984.
43. *Norplant: A Summary of Scientific Data*, The Population Council, New York, 1990.
44. *Contraceptive Technology Update* **12**(1), 1 (1991).
45. E. Diczfalusy, *Contraception* **33**, 7 (1986).
46. *Contraceptive Technology Update* **12**(5), 79 (1991).
47. E. S. E. Hafez, *Obstet. Gynecol. Surv.* **32**, 417 (1977).
48. R. J. Aitken and M. J. K. Harper, *Contraception* **16**, 227 (1977).
49. J. E. Creange, H. P. Schane, A. J. Anzalone, and G. O. Potts, *Fertil. Steril.* **30**, 86 (1978).
50. G. Dipasquale and R. Richter, *Res. Commun. Chem. Pathol. Pharmacol.* **7**, 701, (1974).
51. G. W. Duncan and R. G. Wheeler, *Biol. Reprod.* **12**, 143 (1975).
52. C. W. Emmens, *J. Reprod. Fertil. (Suppl)* **18**, 177 (1973).
53. G. M. Filshie, *Proc. R. Soc. Lond. Biol.* **195**, 115 (1976).
54. J. F. Grunwell, H. D. Benson, J. O'Neal, and V. Petrow, *Steroids* **27**, 759 (1976).
55. D. S. Johnson, *Contraception* **5**, 237 (1972).
56. J. M. Morris and G. Van Wagenen, *Am. J. Obstet. Gynecol.* **115**, 101 (1973).
57. D. W. Hahn, G. Allen, J. L. McGuire, and J. P. Davanzo, *Contraception* **9**, 393 (1974).
58. R. H. Naqvi and J. C. Warren, *Steroids* **18**, 731 (1971).
59. D. M. Potts, *Br. Med. J.* **26**, 65 (1970).
60. L. Schenkel-Hulliger and co-workers, *J. Steroid Biochem.* **11**, 757 (1979).
61. S. J. Segal, *Frontiers in Reproduction and Fertility Control*, MIT Press, Cambridge, 1977, p. 170.
62. *Advances in Methods of Fertility Regulation*, World Health Organization Technical Report Series No. 575, Geneva, 1975.
63. *Out Look* **8**(3), (1990).
64. G. Zuliani, U. F. Colombo, and R. Molla, *European J. OB/GYN and Reprod. Biol.* **37**, 253 (1990).
65. M. Fasoli, F. Parazzini, G. Cecchetti, and C. LaVacchia, *Contraception* **39**, 459 (1990).
66. S. S. Lyer, *Drug News and Perspectives* **4**(4), 228 (1991).
67. V. P. Kamboj, A. B. Kar, S. Roy, and co-workers, *Ind. J. Exp. Biol.* **9**, 103 (1971).
68. *Drugs of the Future* **2**, 441 (1977).

69. C. Tietze and S. K. Henshaw, *Induced Abortion: A World Review*, The Alan Guttmacher Institute, New York and Washington, D.C., 1986, p. 29.
70. G. L. Zatuchini, M. H. Labbok, and J. J. Sciarra eds., *Research Frontiers in Fertility Regulation*, Harper and Row, Hagerstown, Md., 1980, p. 330.
71. *Prostaglandins, Population Reports*, Series E, no. 8, Population Information Program, The Johns Hopkins University, Baltimore, 1980.
72. E. E. Baulieu and S. J. Segal eds., *The Antiprogesterin Steroid RU486 and Human Fertility Control*, Plenum Press, New York, 1985.
73. O. M. Aurech, I. Bukovsky, A. Golan, and co-workers, *Fert. Ster.* **56**, 385 (1991).
74. G. D. Hodgen, *Fert. Ster.* **56**, 394 (1991).
75. W. Elger, W. Beier, K. Chwalisz, and co-workers, *J. Steroid. Biochem.* **25**, 835 (1986).
76. G. I. Zatuchini, A. J. Sobrero, J. J. Speidel, and J. J. Sciarra eds., *Vaginal Contraception*, Harper and Row, Hagerstown, Md., 1979.
77. D. R. Mishell, Jr., *New Engl. J. Med.* **320**, 777 (1989).
78. A. Kolliker, *Z. Wiss. Zool.* **7**, 221 (1855).
79. J. R. Baker, *The Chemical Control of Conception*, Chapman and Hall, London, 1935, p. 151.
80. P. Kestelman and R. L. Kleinman, *Directory of Contraceptives*, International Planned Parenthood Federation (IPPF), London, 1981.
81. R. F. McConnell, in M. H. Briggs and E. Diczfalusy eds., *Pharmacological Models in Contraceptive Development*, WHO Research and Training Center on Human Reproduction, 1974, p. 375.
82. J. R. Baker, *J. Hyg.* **32**, 550 (1932).
83. F. V. Sander and S. D. Cramer, *Human Fertil.*, 6 (1941).
84. **U.S. Pat. 2,752,284** (1956), V. R. Berliner, W. C. Mende, and H. O. Singher.
85. S. Iwahara and K. Furuse, *Jpn. Public Health* **12**, 123 (1965).
86. D. W. Hahn, R. E. Homm, and B. E. McKenzie in G. I. Zatuchni, A. J. Sobrero, J. J. Speidel and J. J. Sciarra eds., *Vaginal Contraception*, Harper and Row, New York, 1980, p. 232.
87. E. B. Connell, in D. R. Mishell, Jr. ed., *Advances in Fertility Research*, Raven Press, New York, 1982, p. 19.
88. E. Diczfalusy, *Contraception* **43**, 201 (1991).
89. R. L. Kleinman, *Intrauterine Contraception*, 4th ed., International Planned Parenthood Federation, London, 1977.
90. F. Alvarez, V. Branche, E. Fernandez, and co-workers, *Fertil. Steril.* **49**, 768 (1988).
91. I. Sivin, *Studies in Family Planning* **20**, 355 (1989).
92. D. R. Mishell, Jr., in J. R. Newton ed., *Clinics in Obstetrics and Gynaecology*, Vol. 11, W. B. Saunders, London, 1984, p. 679.
93. H. J. Tatum, in D. R. Mishell, Jr. ed., *Advances in Fertility Research*, Vol. 1, Raven Press, New York, 1982, p. 47.
94. D. A. Edelman, L. P. Cole, R. Apelo, and P. Lavin, in G. L. Zatuchni, A. Goldsmith, J. D. Shelton and J. J. Sciarra eds., *Long-Acting Contraceptive Delivery Systems*, Harper and Row, Philadelphia, 1984, p. 621.
95. *Intrauterine Devices, Population Reports*, Series B, no. 3, Population Information Program, The Johns Hopkins University, Baltimore, 1979.
96. *Vasectomy: New Opportunities, Population Reports*, Series D, no. 5, Population Information Program, The Johns Hopkins University, Baltimore, 1992.
97. *Voluntary Female Sterilization, Population Reports*, Series C, no. 10, Population Information Program, The Johns Hopkins University, Baltimore, 1990.
98. J. R. Newton, in J. R. Newton, ed., *Clinics in Obstetrics and Gynaecology*, Vol. 11, W. B. Saunders, London, 1984, p. 603.
99. *The Contraceptive Diaphragm*, Emory University, School of Medicine, Healthcare Communications Network, New York, 1989.
100. *Contraceptive Technology Update* **11**(4), 49 (1990).
101. *Condoms, Population Reports*, Series H, no. 8, Population Information Program, The Johns Hopkins University, Baltimore, 1990.
102. *Contraceptive Technology Update* **11**(7), 106 (1990).
103. *Contraceptive Technology Update* **11**(10), 145 (1990).
104. R. L. Kleinman, *Periodic Abstinence for Family Planning*, IPPF, London, 1983.
105. H. Knaus, *Zentralbl. Gynakol.* **53**, 3192 (1929).
106. K. Ogino, *Zentralbl. Gynakol.* **54**, 464 (1930).
107. T. H. van de Velde, *Die Volkommene Ehe, eine Studie uber ihre Physiologie and Technik*, 21st ed., Benno Konegen, Medizinischer Verlag, Leipzig-Stuttgart, 1928.
108. R. J. Woolley, *J. Am. Board Fam. Pract.* **4**, 33 (1991).
109. R. H. Gray, *Lancet*, **335**, 25 (1990).
110. J. Bonnar, in D. R. Mishell, Jr., ed., *Advances in Fertility Research*, Vol. 1, Raven Press, New York, 1982, p. 1.
111. V. C. Stevens, in G. Benagiano and E. Diczfalusy, *Endocrine Mechanisms in Fertility Regulation*, Raven Press, New York, 1983, p. 141.
112. G. P. Talwar, in J. R. Mishell, Jr., ed., *Advances in Fertility Research*, vol. 1, Raven Press, New York, 1982, p. 171.
113. D. J. Anderson and N. J. Alexander, *Fertil. Steril.* **40**, 557 (1982).
114. N. J. Alexander, D. Griffin, J. M. Spieler, and G. M. H. Waites eds., *Gamete Interaction*, Wiley-Liss, New York, 1990.
115. P. D. Griffin, *Human Reprod.* **6**, 166 (1991).
116. G. P. Talwar and A. Gaur, in S. S. Ratnam and E. S. Teoh, eds., *Advances in Fertility and Sterility Series* **6**, 39 (1986).
117. N. C. Sharma, S. Kaul, and L. V. Rao, *Progress in Immunology VI*, Academic Press, New York, 1986, abstract No. 6.23.2.
118. G. P. Talwar, O. Singh, V. Singh, and co-workers, *Fert. Steril.* **46**, 120 (1986).
119. J. C. Herr, R. W. Wright, E. John, and co-workers, in N. J. Alexander, P. D. Griffin, J. M. Spieler and G. M. H. Waites, eds., *Gamete Interaction: Prospects for Immuno-contraception*, Wiley-Liss, New York, 1990, p. 13.
120. P. Primakoff, W. Lathrop, L. Woolman, and co-workers, *Nature* **335**, 543 (1988).
121. E. Goldberg, in N. J. Alexander, P. D. Griffin, J. M. Spieler, and G. M. H. Waites eds., *Gamete Interaction*, Wiley-Liss, New York, 1990, p. 63.
122. T. M. Timmons, S. M. Skinner, and B. S. Dunbar, in N. J. Alexander, P. D. Griffin, J. M. Spieler, and G. M. H. Waites eds., *Gamete Interaction: Prospects for Immunocontraception*, Wiley-Liss, New York, 1990, p. 277.
123. A. V. Schally, in F. Labrie, A. Belanger, and A. Dupont, eds., *LHRH and its Analogues—Basic and Clinical Aspects*, Excerpta Medica, Amsterdam, 1984, p. 3.
124. G. I. Zatuchni, J. D. Shelton, and J. J. Sciarra, eds., *LHRH Peptides as Female and Male Contraceptives*, Harper and Row, Philadelphia, 1981.
125. B. H. Vickery and J. J. Nestor, Jr., eds., *LHRH and its Analogs*, MTP Press, Lancaster, UK, 1987.
126. B. H. Vickery and B. Luneneld, *GnRH Analogs in Cancer and Human Reproduction*, Kluwer Academic, Lancaster, UK, 1989.
127. K. Miyamoto, Y. Hasegawa, M. Fukuda, and co-workers, *Biochem. Biophys. Res. Comm.* **129**, 396 (1985).
128. N. Ling, S. Y. Ying, N. Ueno, and co-workers, *Proc. Natl. Acad. Sci. (USA)*, **82**, 7217 (1985).
129. J. Rivier, J. Speiss, R. McClintock, and co-workers, *Biochem. Biophys. Res. Comm.* **133**, 120 (1985).
130. D. M. Robertson, L. M. Foulds, L. Leversha, and co-workers, *Biochem. Biophys. Res. Comm.* **126**, 220 (1985).
131. N. Ling, S. Y. Ying, N. Ueno, and co-workers, *Nature* **321**, 779 (1986).
132. W. Vale, J. Rivier, J. Vaughan, and co-workers, *Nature* **321**, 776 (1986).
133. H. P. Schane and J. E. Creange, in G. L. Zatuchni, M. Labbok, and J. J. Sciarra, eds., *Research Frontiers in Fertility Regulation*, Harper and Row, Hagerstown, Md., 1980.
134. A. Bartke, D. W. Hahn, R. G. Foldes, and J. L. McGuire, in G. L. Zatuchni, A. Goldsmith, J. M. Spieler, and J. J. Sciarra, eds., *Male Contraception*, Harper and Row, Philadelphia, 1985, p. 158.
135. C. A. Paulsen, W. J. Bremner, and J. M. Leonard, in D. R. Mishell, Jr., ed., *Advances in Fertility Research*, Vol. 1, Raven Press, New York, 1982, p. 157.
136. F. B. Akhtar, G. F. Weinbauer, and E. Nieschlag, *J. Endocrinol.* **104**, 345 (1984).
137. G. F. Weinbauer, F. J. Surmann, F. B. Akhtar, and co-workers, *Fertil. Steril.* **42**, 906 (1984).

138. F. Jockenhovel and co-workers, *J. Clin. Endocrinol. Metab.* **66**, 1065 (1988).
139. G. F. Weinbauer, F. J. Surmann, and E. Nieschlag, *Acta Endocrinol. (Copenhagen)* **114**, 138 (1987).
140. C. J. Bagatell, R. I. McLachlan, D. M. Kretser, and co-workers, *J. Clin. Endocrinol. Metab.* **69**, 43 (1989).

D. W. Hahn
J. L. McGuire
R. W. Johnson Pharmaceutical Research Institute

Gabriel Bialy
National Institutes of Health



CONTROLLED RELEASE TECHNOLOGY

Agricultural,
Pharmaceutical,

AGRICULTURAL

Controlled release technology in agriculture encompasses the controlled delivery of plant nutrients, ie, fertilizers, as well as control chemicals, eg, herbicides (qv), insecticides, fungicides (qv), etc, to a target in a manner which maximizes its use efficiency, minimizes potential negative effects associated with overdosage, and or extends the time in which sufficient dosages are delivered. Controlled release fertilizer (CRF) technologies are emphasized here.

A goal of controlled release fertilizer research since the 1940s has been the development of a product that delivers its nutrients at a rate matching the demand rate of the plant to which it is applied. Such a fertilizer would represent the ultimate in use efficiency; agronomic performance, ie, crop yield, quality, and appearance; agronomic safety; and labor savings, ie, reduced application frequency. It also would minimize potential losses to the environment.

The benefits of controlled release fertilizers come at a cost premium, eg, the cost of controlled release nitrogen (CRN) can be anywhere from 2.5 to 10 times the cost of nitrogen from a conventional fertilizer such as urea. Because of the large price differential with conventional fertilizers, usage of CRFs has been limited. In 1990, controlled release nitrogen sources accounted for only 1% of total U.S. nitrogen fertilizer consumption. Usage of CRFs has been limited primarily to specialty markets where their advantages justify the increased cost, ie, high to medium value agricultural crops, consumer lawn and garden, lawn and landscape care companies, golf courses, nurseries and greenhouses, and professional turf applications.

The wholesale value of CRFs serving the U.S. specialty markets was estimated at \$175 million in 1990. Of this total, agriculture comprised \$14 million (6.8 thousand t of CRN); nonagricultural markets accounted for \$161 million (93.9 thousand t of CRN).

U.S. agricultural use of controlled release fertilizers is concentrated primarily in the Southeast and California on high value crops such as strawberries, tomatoes, bell peppers, melons, and citrus. Figure 1 shows the breakdown of agricultural use of CRFs both by crop type and product. Strawberry production accounts for over half of the 1990 agricultural consumption of CRFs in dollars. However sulfur-coated urea (SCU) accounts for the largest tonnage (61%) of agricultural usage.

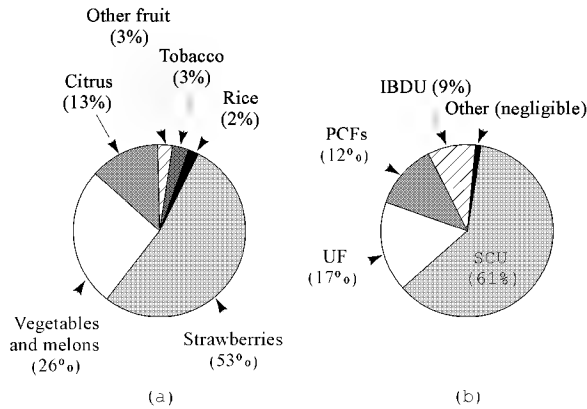


Fig. 1. United States agricultural crop markets for manufactured controlled release fertilizers, 1990. (a) Total percentage (\$14 million) of market value by crop; and (b) total percentage (6,800 t) of consumption by product type. IBDU is isobutylidene diurea; PCF, polymer-coated fertilizer; UF, urea-formaldehyde; SCU, sulfur-coated urea.

Courtesy of SRI International.

Nonagricultural markets represented significant outlets for CRFs in the United States in 1990. In total, these markets accounted for 93% of the demand for CRN. As shown in Figure 2, about 3-4 of the nonagricultural market (\$161 million total) is divided approximately evenly between golf course, consumer lawn and garden, and nurseries and greenhouses. Urea-formaldehyde (UF) products were the largest source of CRN for these markets (63% of total), followed by SCU (28% of total).

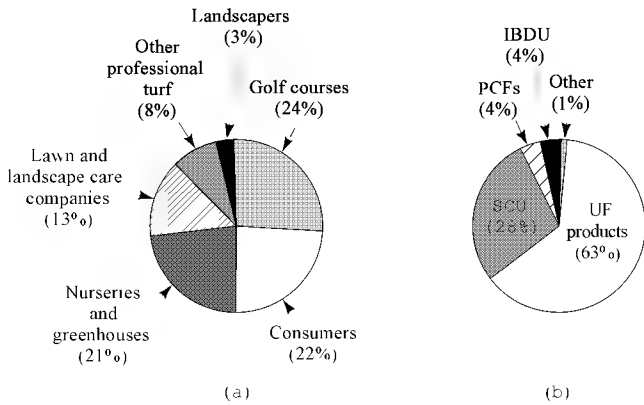


Fig. 2. United States nonagricultural markets for manufactured controlled release fertilizers, 1990. (a) Total percentage (\$161 million) of nonagricultural market value; and (b) total percentage (93,900 t) of controlled release nitrogen consumption by product type. IBDU is isobutylidene diurea; PCF, polymer-coated fertilizer; UF, urea-formaldehyde; SCU, sulfur-coated urea.

Courtesy of SRI International.

Numerous technologies have been developed with varying degrees of success to maximize the various benefits of controlled release. Heightened attention to environmental concerns has prompted increased activities worldwide in the development and commercialization of controlled release fertilizers.

Synthetically produced CRFs are categorized either as nitrogen reaction products or coated fertilizers. Nitrogen reaction products are produced by the chemical reaction of water-soluble nitrogen compounds, such as urea or ammonia, to create nitrogen fertilizers possessing more complex molecular structures. These complex molecules have limited water solubility and are slowly broken down in the soil environment to nitrogen forms which the plant can assimilate. The limited solubility and the rate of decomposition controls the nitrogen availability to the plant.

Coated fertilizers achieve controlled release by coating a soluble fertilizer core (substrate) with a water-insoluble barrier which limits the access of water to the fertilizer and thus limits its dissolution rate.

Nitrogen Reaction Products

Urea–Formaldehyde Reaction Products. Urea–formaldehyde (UF) reaction products represent one of the older controlled release nitrogen technologies. An early disclosure of the reaction products of urea [57-13-6] and formaldehyde [50-00-0] was made in 1936 (1) (AMINO RESINS AND PLASTICS). In 1948, the USDA reported that urea (qv) and formaldehyde (qv) could react to produce a controlled release fertilizer at urea to formaldehyde mole ratios (UF ratio) greater than one (2).

Urea–formaldehyde reaction products represented the first synthetically produced form of controlled release nitrogen and were commercialized in 1955 under the trade names Uramite (DuPont) and Nitroform (Nitroform Corp.).

The urea–formaldehyde reaction results in a distribution of methylene urea (MU) polymers of varying molecular weights or polymer chain lengths and of varying water solubility.

Granular Compositions. Granular compositions of UF reaction products are divided into three classes, each based on their degree of water solubility as affected by their polymer distributions, ie, ureaform, the least water-soluble class; methylene ureas; and methylene diurea dimethylene triurea (MDU DMTU) compositions, the shortest-chain MU oligomers and the most water-soluble.

Each class of UF products contains certain amounts of unreacted urea. MDU DMTU compositions, the lowest molecular-weight distribution, contain the highest amount of unreacted urea nitrogen, and ureaform contain the least amount.

Ureaform is the oldest class of UF reaction products. Ureaform, as defined by the American Association of Plant Food Control Official (AAPFCO-1964), is sparingly soluble. It contains at least 35% total nitrogen with at least 60% of the total nitrogen as cold water-insoluble nitrogen (CWIN), determined by AOAC Method 945.01 (1990). Further, it must have an Activity Index (AI), ie, the percent of CWIN that is soluble in hot (100°C) water, of not less than 40%. Ureaform is composed largely of longer-chained UF polymers, primarily tetramethylene pentaurea (TMPU) and longer. Unreacted urea nitrogen content is usually less than 15% of the total nitrogen.

Methylene ureas are a class of sparingly soluble products which evolved during 1960s and 1970s. These products contain predominantly intermediate chain-length polymers, primarily trimethylene tetraurea (TMTU) and tetramethylene pentaurea (TMPU). The total nitrogen content of these polymers is 39 to 40% with between 25 and 60% of the nitrogen present as CWIN. The unreacted urea content generally is in the range of 15 to 30% of the total nitrogen. This class of granular UF products is not specifically defined under AAPFCO Official Terms and Definitions.

MDU DMTU compositions are the newest class of UF fertilizers and were developed during the 1980s. These granular compositions are predominantly shorter-chain polymers with at least 60% of their polymeric nitrogen in the form of the cold water-soluble polymers methylene diurea (MDU) and dimethylene triurea (DMTU). The total nitrogen content of these polymers is above 40%; generally less than 25% of their total nitrogen is present as CWIN. Because of the shift in polymer distribution to lower-weight polymers, these compositions contain higher percentages of unreacted urea than ureaforms and methylene ureas. AAPFCO defines MDU and DMTU individually as sources of slowly available nitrogen.

Liquid Compositions. Urea–formaldehyde reaction products also are available commercially as liquids that can be categorized into two classes, ie, water suspensions and water solutions.

UF suspensions comprise a family of liquid products in which the UF reaction products are present as a microcrystalline dispersion of longer-chain UF polymers (CWIN) in a water solution of urea and water-soluble UF compounds. Typically, these products contain about 25% of the total nitrogen as CWIN.

UF solutions are clear water solutions. They contain only very low molecular-weight, water-soluble UF reaction products plus unreacted urea. Various combinations of UF solutions are found. They contain a maximum of 55% unreacted urea with the remainder as one or more of methylolureas, methylolurea ethers, MDU, DMTU, or triazone, a cyclical oligomer. AAPFCO has defined this class of compounds as urea–formaldehyde products (water-soluble).

Physical and Chemical Properties. The reaction of urea and formaldehyde forms a white solid. The solubility varies with the methylene urea polymer chain length; longer-chain, higher molecular-weight UF polymers are less water-soluble than short-chain polymers. Physical properties of the methylene urea polymers which have been isolated are compared to urea in Table 1.

Table 1. Physical Properties of Methylene Urea Polymers

Product	Melting point, °C	Water solubility, g 100 g	Temperature, °C
urea	132.7	100	17
MDU	205–207	2.5	25
		7.0	50
		0.1	25
		4.4	100
DMTU	231–232		

Prior to the development of analytical techniques to quantify specific methylene urea oligomers, methylene urea polymer distributions were characterized by physical (solubility) methods. Products were separated into three fractions (3).

Fraction I is cold water-soluble nitrogen (CWSN), ie, that portion of the total nitrogen that is soluble in cold water (22°C) when 1 or 1.4 g of product is placed in 250 mL of water. Fraction I can consist of unreacted or residual urea, methylol ureas, MDU, DMTU, and other soluble UF reaction products. Fraction II is the cold (22°C) water-insoluble nitrogen (CWIN) which is hot (100°C) water-soluble (HWS). Compositionally, this fraction comprises the medium molecular-weight polymers, TMTU and TMPU. Fraction III is hot (100°C) water-insoluble nitrogen (HWIN). Compositionally, these are the higher molecular-weight methylene urea polymers, TMPU and higher.

The nitrogen content of granular urea–formaldehyde reaction products typically ranges from 35 to 42% depending on the methylene urea polymer distribution.

An improved method (4) of characterizing the lower molecular-weight, ie, water-soluble methylene ureas, was developed in the 1980s using high pressure liquid chromatography (hplc). This method permits the identification of urea and individual water-soluble methylene urea oligomers, and the characterization of their distribution. By this method, urea, MDU, and DMTU can be quantified; development of a similar method for quantifying TMTU and TMPU is underway. Higher molecular-weight polymers continue to be characterized by solubilities in cold and hot water, ie, CWIN and HWIN.

Table 2 gives the typical total nitrogen analysis and the unreacted urea and UF polymer distribution of various UF reaction products.

Table 2. Total Nitrogen Analysis of UF Reaction Products

Total nitrogen as specific fraction ^a , %
--

UF reaction product	Total nitrogen, %	Fraction I, % composition ^b	Fraction II, % composition ^c	Fraction III, % composition ^d
<i>Granular products</i>				
ureaform	35–38	30 urea, MDU, DMTU	40 TMTU, TMPU	30
methylene ureas	39–40	65 urea, MDU, DMTU	20 TMTU, DMTU	15
MDU DMTU	40–42	90 urea, MDU, DMTU	7 TMTU	3
<i>Liquid products</i>				
UF suspension compositions	18	75 urea, methylol ureas, MDU, DMTU	15 TMTU, TMPU	10
methylol urea solutions	26–30	100 urea, methylol ureas		
MDU solutions	26	100 urea, methylol ureas, MDU		
urea–triazone solutions	28	100 urea, triazone		

^a Fraction I: cold water-soluble nitrogen (CWSN); Fraction II: cold water-insoluble nitrogen (CWIN), which is hot water-soluble; Fraction III: hot water-insoluble nitrogen (HWIN).

^b MDU: methylene diurea [13547-17-6]; DMTU: dimethylene triurea [15499-91-9].

^c TMTU: trimethylene tetraurea [35710-95-3]; TMPU: tetramethylene pentaurea [50837-30-4].

^d Compositions is mix of pentamethylene hexaurea (PMHU) [35710-96-4] and higher molecular-weight urea.

Agronomic Properties and Nutrient Release Mechanism. The conversion of UF reaction products to plant available nitrogen is a multistep process, involving dissolution and decomposition. Materials are slow to enter the soil solution by virtue of their low solubility. Longer polymer chain products are less soluble than shorter chains and take longer to become available to the plants.

Once in the soil solution, urea–formaldehyde reaction products are converted to plant available nitrogen through either microbial decomposition or hydrolysis. Microbial decomposition is the primary mechanism. The carbon in the methylene urea polymers is the site of microbial activity. Environmental factors that affect soil microbial activity also affect the nitrogen availability of UF products. These factors include soil temperature, moisture, pH, and aeration or oxygen availability.

The rate of nitrogen released (mineralization) from UF reaction products is directly affected by polymer chain length. The longer the methylene urea polymer length, the longer it takes for the nitrogen to become fully available. For ureaform and methylene urea products, the rate of mineralization is reflected by the cold water insoluble nitrogen (CWIN) content and its Activity Index. Figure 3 shows the results of a greenhouse study in which Kentucky bluegrass was treated with 2 lb N / 1000 sq ft (2 lb N / M) (9.76 g / m²) of a number of UF products varying in CWIN content. As shown, the rate of nitrogen availability, as reflected by the leaf fresh weight, was directly correlated to the CWIN content. It should be noted that the efficiency of nitrogen utilization also decreases with increasing polymer length, as reflected by a decrease in the Activity Index. It is questionable if the very long methylene urea polymers (HWIN) are effectively used by the crop.

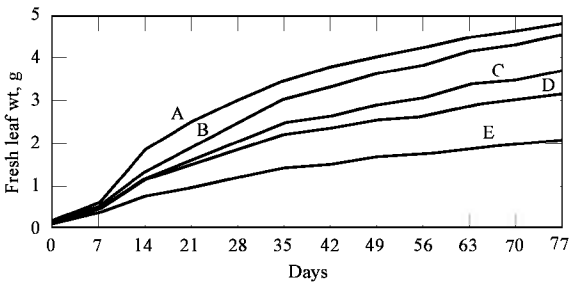
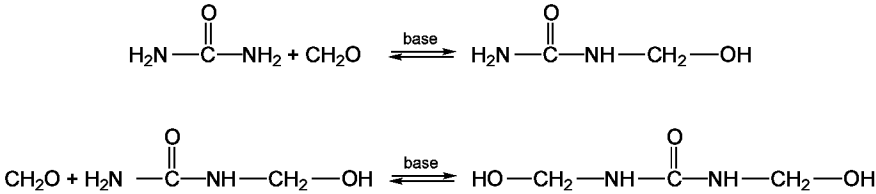


Fig. 3. Response of Kentucky bluegrass to urea, A, and methylene urea (MU) products varying in cold water-insoluble nitrogen (CWIN). B is 27% CWIN; C, 50% CWIN; D, 65% CWIN; and E, 80% CWIN. Application of 2 lb N / 1000 sq ft (9.76 g / m²).

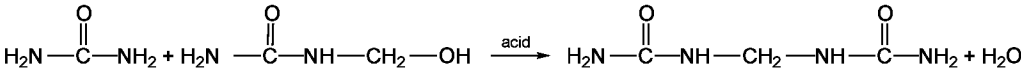
Courtesy of The O.M. Scott & Sons, Co.

Nitrogen release from MDU / DMTU compositions cannot be predicted from WIN values, since these polymers are largely water-soluble. Instead, hplc procedures must be used to characterize these products. MDU / DMTU compositions were developed to maximize the nitrogen efficiency of UF reaction products. They have greater agronomic efficiency than products containing high levels of water-insoluble polymers and yet retain agronomic and environmental safety over conventional soluble fertilizers such as urea by nature of their greatly reduced solubility.

Manufacturing and Processing. Urea–formaldehyde fertilizers are prepared from the reaction of formaldehyde with excess urea (U / F mole ratio 1). The reaction proceeds in two steps. In Step 1, urea and formaldehyde are combined to form monomethylol [1000-82-4] and dimethylolurea [140-95-4] intermediates.



These compounds are stable only under basic conditions and must be held at high pH if stored for considerable periods of time. In Step 2, these methylolureas further react under acidic conditions with urea to form the various methylene urea oligomers. Water generated from the condensation reaction is removed by evaporation when producing granular UF products.



Methylene diurea (MDU) is the shortest chain of the urea–formaldehyde oligomers. MDU can further react with additional monomethylol urea to form dimethylene triurea (DMTU), the next higher oligomer. DMTU can also be formed from the condensation reaction of urea with dimethylol urea. In similar fashion, higher chain-length polymers are obtained by continued reaction with additional methylolureas.

In actual practice, the reaction of urea with formaldehyde produces a distribution of polymers of varying chain length. The distribution is affected by the U / F mole ratio as well as reaction conditions such as pH, temperature, and reaction time. In general, higher U / F ratios produce polymer distributions

water solution	84		
solid urea liquid	54		
aldehyde			
specific gravity			
true	1.3	1.7	1.6
apparent	0.7–0.9		
sublimation temperature, °C		265	
decomposition temperature, °C		290	345

^a Commercial-grade product. Theoretical content is 32.18%.

^b Fertilizer-grade product.

^c High nitrogen content distinguishes it from other controlled release nitrogen sources.

^d With decomposition.

^e To convert J to cal, divide by 4.184.

Isobutylidene diurea [6104-30-9] is a nonhygroscopic white crystalline solid available in fine (0.5–1.0 mm), coarse (0.7–2.5 mm), and chunk (2.0–3.0 mm) particle sizes. The AAPFCO official definition requires a minimum nitrogen content of 30% with 90% of the nitrogen in water-insoluble form prior to grinding (3).

Nitrogen from isobutylidene diurea becomes available to plants through hydrolysis. In the presence of water, the compound will hydrolyze to urea and isobutyraldehyde. The rate of hydrolysis is accelerated by low pH and high temperatures. Unlike UF polymers that rely on soil microbial populations to make the nitrogen available, isobutylidene diurea is primarily dependent on water as the critical element to nitrogen availability. Its low water solubility controls the transport of the product into the soil solution. Once in the soil solution, the rate of hydrolysis is affected by both soil pH and temperature. Because of its limited solubility, the rate of dissolution is affected by the particle size of the product and the amount of water available. The powder form is mineralized much more rapidly than large particle granules under the same field conditions. Because the release is not microbe dependent, it can become available at low temperatures. This and the dependency on moisture are the distinguishing characteristics of isobutylidene diurea.

The agronomic response of IBDU is compared to other nitrogen reaction products in Figure 4. In this study, various products were applied to Kentucky bluegrass at 3 lb N M (14.6 g m²). Plant growth as measured by leaf fresh weight was monitored weekly.

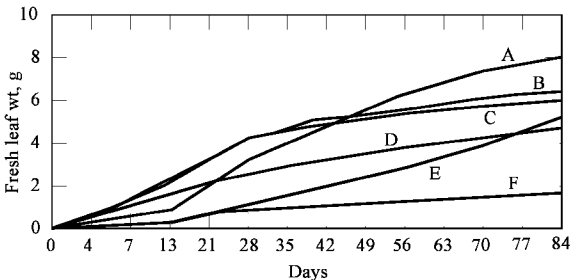


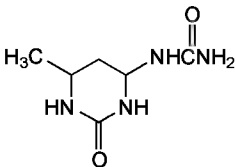
Fig. 4. Response of Kentucky bluegrass to select nitrogen reaction controlled release products. Application of 3 lbs N 1000 ft² (14.6 g m²). A is IBDU; B, urea; C, oxamide (powder); D, nitroform; E, oxamide (granular); and F, melamine.

Courtesy of The O.M. Scott & Sons Co.

Manufacturing, Processing, and Uses. Isobutylidene diurea is manufactured by the condensation reaction of urea and isobutyraldehyde. The reaction can be carried out both in aqueous solution and between solid urea and the liquid aldehyde. The presence of acid catalyzes the reaction. Commercially, isobutylidene diurea is made by spraying the liquid aldehyde into a paddle mixer containing acid treated solid urea. Patents on the technology are held by Mitsubishi Chemical Industries (Japan) and Farbwerk Hoechst Aktiengesellschaft (Germany). Mitsubishi produces IBDU from isobutyraldehyde recovered as a by-product of 2-ethylhexyl alcohol production. Prior to 1986, all isobutylidene diurea consumed in the United States was imported from Japan. In 1986, IB Chemicals brought domestic production on stream with a 20,000 t yr facility in Bucks, Alabama. IB Chemicals is a joint venture company held by Virginia Chemicals, Inc. (Hoechst Celanese Corp.), Mitsubishi Chemical Industries of America, and Mitsubishi International Corp. Briquetted forms, ie, Woodace Briquettes, Mitsubishi Chemical Industries, continue to be imported from Japan.

IBDU is used on turfgrasses, in commercial nurseries, and in landscaping, forestry, and specialty (medium to high value) agriculture. Although some fine-size IBDU 31-0-0 is used for direct application to golf course greens, most of the turfgrass use is in the form of blended fertilizers, often in combination with other types of controlled release fertilizers (CRF). Vigoro Industries has exclusive marketing rights in North America and sells to turfgrass markets under the trade name ParEx, to nurseries as Woodace, and to specialty agriculture under the name Certified Harvest King.

Crotonylidene Diurea (CDU). Crotonylidene [1129-42-6] is produced by the acid catalyzed reaction of urea with either crotonaldehyde or acetaldehyde. The condensation reaction produces a ring-structured compound. Table 4 lists select properties.



CDU in pure form is a white powder. It is made slowly available to the soil solution by nature of its limited solubility in water. Once in the soil solution, nitrogen from CDU is made available to the plant through a combination of hydrolysis and microbial decomposition. As with any CRF which is dependent on microbial action, the mineralization of CDU is temperature dependent. Product particle size has a significant effect on CDU nitrogen release rate. Smaller particles mineralize more rapidly because of the larger surface contact with the soil solution and the microbial environment. The rate of nitrogen release is also affected by pH because CDU degrades more rapidly in acidic soils.

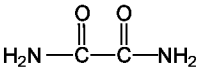
Manufacturing, Processing, and Uses. In commercial production, aqueous urea solution is mixed with acetaldehyde in 1:1 molar ratios. An acid catalyst is introduced into the reaction mixture which is staged at various process temperatures. After neutralization with a base, the CDU is separated from the mother liquor by filtration or spray drying.

CDU is manufactured by Chisso Corp. (Japan) and marketed as CDU by Chisso Asahi Fertilizer Co., Ltd. It is also produced by BASF (Germany) under Japanese license. It is sold under the trade name Crotodur in Europe.

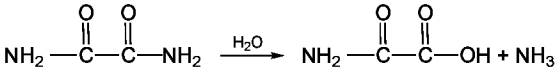
CDU has very limited use in the United States. It is used primarily in Japan and Europe where it is produced. It serves the turf and specialty agriculture markets and is typically formulated into granulated N–P–K fertilizers.

Other Reaction Products. Nitrogen reaction fertilizers are commercially available that do not involve reactions between urea and aldehydes. These are oxamide and melamine.

Oxamide. Oxamide [471-46-5] is a nonhygroscopic single compound. It has a molecular weight of 88.08, a nitrogen content of 31.8%, and is a white crystalline solid with very limited solubility in water. Table 4 lists select physical properties.



Following slow dissolution into the soil solution, oxamide undergoes stepwise hydrolysis to liberate ammonia. Oxamic acid is formed in the first step.



Oxalic acid, HOOC-COOH, the product of the second hydrolysis step, can be toxic to plants if present in sufficient quantity before being further converted by soil microbes to carbon dioxide. As with other insoluble products that release nitrogen through hydrolysis, eg, IBDU, the rate of hydrolysis is greatly affected by the particle size of the product which affects dissolution rate into the soil solution. Unlike isobutylidene diurea, however, the release rate is not affected by pH. Oxamide (granular) is a very slow release nitrogen source. The agronomic activity of oxamide (granular and powder) on Kentucky bluegrass is compared to other nitrogen reaction products sources in Figure 4.

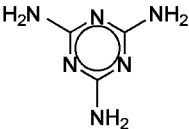
Manufacturing, Processing, and Uses. Oxamide can be prepared by several methods. One method involves the oxidation of hydrogen cyanide followed by hydrolysis to oxamide by a two-step reaction. Hydrogen cyanide (HCN) is oxidized with either copper oxide or nitrogen dioxide to produce cyanogen (C₂N₂). Cyanogen is then converted to oxamide by acid-catalyzed hydrolysis using concentrated hydrochloric acid or a mixture of hydrochloric and acetic acids.

Another method, utilized by Ube Industries, Ltd. (Japan), involves the oxidation of methanol and carbon monoxide to form an oxalic diester. This then reacts with ammonia in a second step to form oxamide.

Oxamide is produced commercially by Ube Industries, Ltd. (Japan) and a pilot process is being operated by Enichem (Italy). It is not produced domestically as a commercial fertilizer, although it was the subject of much research and development activity by the Tennessee Valley Authority's (TVA) National Fertilizer Research and Development Center. It is made in small quantities for industrial use by Allied Chemical, Hummel Chemical Co., and United Guardian, Inc. Oxamide has application as a controlled release nitrogen source for the turf and specialty agricultural markets.

Melamine. Melamine, 2,4,6-triamino-1,3,5-triazine [108-78-1] is a white crystalline powder, sparingly soluble in cold and hot water, with a molecular weight of 126 (Table 4). It has a very slow mineralization rate. After an initial delay of one to two months, it releases very slowly by microbial decomposition over the period of one to two years. Because of this slow availability, melamine has been cogranulated with urea to increase the upfront nitrogen response. The agronomic response of melamine on Kentucky bluegrass (greenhouse) is compared to other noncoated controlled release nitrogens in Figure 4.

Melamine is produced by heating urea under pressure in the presence of a catalyst. The result is a ring structure as shown below. The reaction by-products, ammonia and carbon dioxide, can be recycled for urea production.



The production process was developed by Melamine Chemicals, Inc. in the early 1970s. Marketing rights for melamine as a CRF were acquired by Pursell Industries in 1988. Melamine CRF products are currently available as Nitrazine 66N, a melamine powder (66% nitrogen). This powder also can be an agglomeration granulated with molten urea to form a homogeneous granular product containing from 50 to 67% melamine and from 50 to 33% urea. This granular product, Nitrazine-U, which contains 55 to 60% nitrogen, is not currently available commercially.

The slow mineralization rate suggests the use of melamine for very long-term crops. Powdered melamine (Nitrazine 66N) is used in fertilizer spikes and stakes for houseplants and ornamentals.

Coated Fertilizers

Coated fertilizers is a term characterizing those products in which a soluble fertilizer core (substrate) granule is covered with a water-insoluble coating. The coating limits or controls the rate of water penetration to the soluble fertilizer core, and in some products controls the release rate of solubilized fertilizer from within the granule to the external environment. The three categories of coated fertilizers are those using sulfur as the coating material, those that employ a polymeric material, and those hybrid products that utilize a multilayer coating of sulfur and polymer.

Coated fertilizers represent the fastest growing segment in controlled release fertilizer technology. During the 1980s, coated fertilizer products grew at a rate of about 10% per year, UF product usage grew at about 2% per year, and all manufactured controlled release fertilizers, coated and nitrogen reaction products, grew at a rate of about 4% per year. The growth of coated fertilizers is the result of more favorable economics, increased flexibility in nutrient release patterns as compared to nitrogen reaction products, and the flexibility in controlling the release of other nutrients in addition to nitrogen.

Sulfur-Coated Fertilizers. Sulfur-coated urea technology (SCU) was developed in the 1960s and 1970s by the Tennessee Valley Authority, now called the National Fertilizer and Environmental Research Center. A commercial-scale demonstration plant (9.1 t h) was put in operation by TVA in late 1978. Sulfur was chosen as the principle coating material because of its low cost and its value as a secondary nutrient.

Physical and Chemical Properties. Sulfur-coated ureas (SCUs) typically have bulk densities of 720–800 kg m³ (45–50 lb ft³). The color ranges from brown to tan to bright yellow depending on the source of urea, whether or not a sealant is used, and the type sealant employed. Soft wax sealants are typically used as a secondary coating over the sulfur coating to fill in imperfections in the sulfur coating and to provide handling integrity to the brittle sulfur coat.

Product particle sizes vary from standard size of 6–14 mesh–U.S. Std. Sieves to mini-size granules of 10–16 mesh–U.S. Std. Sieves for small particle blends, to micro-size granules of 14–35–U.S. Std. Sieves for use on golf course tees and greens. Approximate mm corresponding to mesh sizes are 6 mesh 3.36 mm; 10 mesh 2 mm; 35 mesh 0.5 mm.

The total nitrogen content of SCUs vary with the amount of coating applied; SCUs available in the early 1990s range from 30 to 40% nitrogen. The release characteristics of SCU are determined by the Katz Method, a 2-h dissolution (WIN) test at room temperature, and adaptations of the 7-day dissolution rate (DR) developed by TVA (9). The Katz Method procedure is similar to the method used to analyze UF reaction products, however, the grinding step is by-passed as this would destroy the sulfur coating. The 7-day DR test involves total immersion of the product in water for 7 days and measurement of the percent of the total nitrogen leached into the water. By pulling aliquots at various time intervals and measuring the amount of nitrogen dissolved, a leach profile can be generated which gives improved insight into the release characteristics of the product. Laboratory WIN (Katz) and 7-day DR tests are reasonably good quality control tools for a given process, but have been found to be inadequate predictors of agronomic response when comparing different technologies. For instance, SCUs with and without wax sealants have been found to have dramatically different agronomic release at similar DRs or Katz WIN values.

Agronomic Properties and Nutrient Release Mechanisms. The mechanism of nutrient release from SCU is by water penetration through micropores and imperfections, ie, cracks or incomplete sulfur coverage, in the coating. This is followed by a rapid release of the dissolved urea from the core of the particle. When wax sealants are used, a dual release mechanism is created. Microbes in the soil environment must attack the sealant to reveal the imperfections in the sulfur coating. Because microbial populations vary with temperature, the release properties of wax-sealed SCUs are also temperature dependent.

The release rate of an SCU particle is directly affected by the coating thickness and the coating quality. Particles with higher sulfur loads, ie, thick sulfur coatings, typically show fewer imperfections than particles with lighter coatings. There is the risk, however, that particles with too thick sulfur coating will exhibit lock-off, ie, they may never effectively release their nutrient. Because of the continuous horizontal rotary coating drum method of coating application, SCU product is in reality a statistical blend of granules with many different coating weights, thicknesses, and qualities. This blend of granules minimizes the effects of individual particles receiving light or heavy coatings.

In turf applications, SCU, with or without wax sealant, provides improved safety and longevity over urea. This is shown in Figure 5 which rates the quality of Kentucky bluegrass with time after being treated with 2 lb N 1000 sq ft (9.76 g m²) of urea and two types of SCU. In this particular test, a quality rating of 3 or less is acceptable. Urea provided acceptable turf quality for 11 weeks, whereas the two SCU products (37% nitrogen) gave quality turf for 16 weeks.

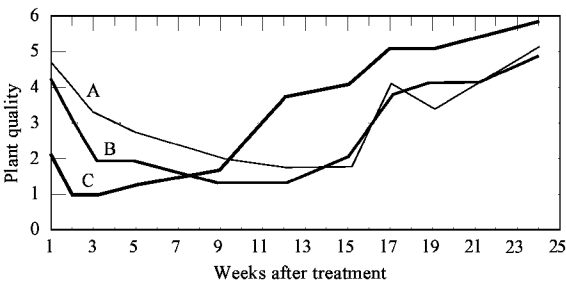


Fig. 5. Response of Kentucky bluegrass to sulfur-coated urea. Application of 2 lbs N 1000 ft² (9.76 g m²). Quality ratings of 1–3 are good, 4–6 fair. Quality rating of 3 is the minimum quality acceptable. A is SCU (wax); B, SCU (no wax); and C, urea.

Courtesy The O.M. Scott & Sons Co.

Depending on the coating weight, nitrogen application rate, and environmental conditions, SCUs can have residual characteristics which provide agronomic response from 6 to 16 weeks in turfgrass applications.

Manufacturing and Processing. The process for sulfur-coated urea, as developed by TVA (10), involves the precision application of three coating materials to a soluble fertilizer core, usually granular urea. The coatings include molten sulfur, a hydrocarbon oil (soft wax) sealant, and a flow conditioner. The process involves preheating urea prior to coating with molten sulfur. Preheating the urea allows better adhesion of the sulfur and more time for the molten sulfur to wrap the urea prior to solidifying. The quality of the urea substrate as well as the sulfur are key to producing a quality product.

Sulfur has two crystalline forms and undergoes a transformation in crystalline structure with time (aging). At the time of application, most of the sulfur in the coating is present as monoclinic sulfur with a small portion of amorphous sulfur. With time, however, the monoclinic sulfur is converted to orthorhombic sulfur which has a slightly higher density. This density change causes shrinkage in the sulfur coating. Because solid sulfur is inelastic and brittle, the shrinkage often causes stress fractures and micropores in the coating. The degree of imperfections can be influenced by the conditions under which the sulfur is applied and the conditions under which the product is cooled. The TVA process employs a soft, waxlike, hydrocarbon oil sealant to effectively seal up any imperfections in the sulfur coating. The sealant most commonly used by industry consists of a 70/30 blend of heavy paraffinic oil, eg, Shell-Flex 790, with a polyethylene wax, eg, Allied AC-6. This sealant is semiflowable at ambient temperature which allows it to fill in cracks as they occur. The soft nature of the wax necessitates application of a flow conditioner, such as diatomaceous earth, to keep the granules from sticking together. Typical SCU compositions of the TVA type consist of 14% sulfur, 2.1% sealant, and 2.5% conditioner.

Some SCU products are produced without sealant. These are produced under carefully controlled process conditions that have been optimized to minimize the formation of stress fractures in the sulfur coating (11). However, such products are more prone to attrition damage when handled than SCU products with a sealant coating.

Sulfur-coated urea is produced in North America by a number of suppliers including ICI Canada (Courtright, Ontario), Lescro, Inc. (Martins Ferry, Ohio), Pursell Industries (Sylacauga, Alabama), and The O.M. Scott & Sons Co. (Marysville, Ohio). All produce a number of fertilizer grades and particle sizes. Sulfur-coated ureas and compound (N–P–K) fertilizers also are produced by Mitsui Toatsu (Japan).

SCU has found use in the following markets. Products serving these markets are generally blends of SCU with other fertilizer components.

Granular particle size	Market
standard	turf (golf courses) lawn-care companies (homeowners) nurseries
mini	specialty agriculture turf (golf tees and greens) lawn-care companies
micro	turf (golf greens)

Polymer-Coated Fertilizers. Polymer-coated fertilizers (PFC) represent the most technically advanced state of the art in terms of controlling product longevity and providing nutrient efficiency. Because most polymer-coated products release by diffusion through a semipermeable membrane, the rate of release can be altered by composition of the coating and the coating thickness. Polymer coatings can be categorized as either thermoset resins or thermoplastic resins (see COMPOSITE MATERIALS, POLYMER MATRIX). Because of the relative high coatings cost of most polymer-coated products compared to SCU, their use has been primarily restricted to high value applications. Polymer-coated fertilizer technologies vary greatly between suppliers.

Grace Sierra Horticultural Products Co. This company uses technology originally patented by Archer Daniels Midland Corp. for the production of polymer-coated products. The alkyd resin coating technology, developed in the early 1960s, involves coating a soluble fertilizer core with a thermoset copolymer of dicyclopentadiene and a glycerol ester (linseed oil) dissolved in an aliphatic hydrocarbon solvent.

Nutrient release patterns vary with the amount of coating applied and the substrate used. Coating weights vary from 10 to 20%. Typically, commercial products are blends of different coating weights. Coated substrates include, but are not limited to, urea, potassium sulfate, and ammonium nitrate-based N–P–K fertilizers.

The Osmocote product line is based on coating prilled N–P–K fertilizers. Product longevities range from 5 to 16 months, depending on the temperature. The Osmocote line also includes a coated N–P–K miniprill which lasts 2–3 months.

The Sierra product line contains secondary and minor elements. Sierrablen products, ie, blends of Osmocote with uncoated nutrients for quicker

initial response, contain iron. Blends of Osmocote with uncoated granules are sold to the specialty agriculture high value crops markets under the trade name Agriform.

High-N products are blends of Osmocote with coated granular urea. These blends are made with heavier coating weights and provide the longest longevities in Grace-Sierra's product line. The cost of the heavier coatings is offset through the use of coated urea which has a lower cost per unit of nitrogen than coated N–P–K complex fertilizers.

Grace-Sierra products, like most polymer-coated products, release by diffusion kinetics through a semipermeable membrane. Water vapor penetrates the resin coating and dissolves the water-soluble fertilizer core. The dissolved nutrients then diffuse back out through the coating to the environment. The release pattern is much more linear than SCU technology, which releases by matrix kinetics through coating imperfections. The difference in release mechanism is easily seen by looking at the growth response of Kentucky bluegrass. Bluegrass treated (2 lb N / 1000 sq ft) (9.76 g / m²) in the greenhouse with a number of polymer-coated fertilizers is compared to urea. Fresh weights were recorded weekly and plotted in Figure 6 as cumulative yield as a function of time. The linear release patterns of the polymer-coated fertilizers are quite apparent.

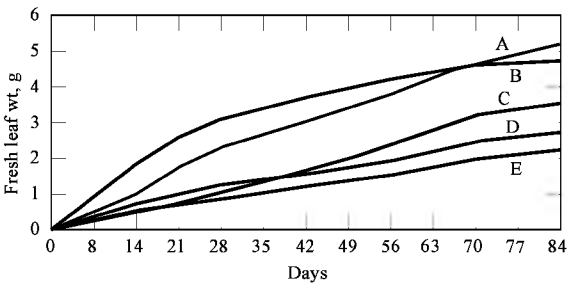


Fig. 6. Response of Kentucky bluegrass to polymer-coated fertilizers of select manufacturers. Application of 2 lbs N / 1000 ft² (9.76 g / m²). A is Multicote 4 (29-0-0) (Haifa); B, urea; C, Pursell; D, Nutricote (Chisso-Asahi); and E, Grace-Sierra.

Courtesy of The O.M. Scott & Sons.

The coating material is a copolymer of dicyclopentadiene and a glycerol ester of linseed oil. This material is dissolved in mineral spirits and applied in multiple layers to the fertilizer in a batch coating drum operating at 65–70°C. About 3% resin is applied in each coating cycle which takes about 20 min. About half the cycle is devoted to applying the coating; the remainder is used for evaporation of the solvent. The solvent is recovered and recycled back into the system. The process goes through repeated coating cycles until the desired coating weight is achieved. Grace-Sierra operates production facilities in Milpitas, California; Charleston, South Carolina; and Heerde, The Netherlands.

High product costs limit distribution to high value crop markets. This includes commercial ornamental production such as nurseries and greenhouses, citrus production, and strawberry production. Limited amounts are sold to the consumer lawn and garden market.

Chisso-Asahi Fertilizer Company, Ltd. This company utilizes thermoplastic resins, such as polyolefins, poly(vinylidene chloride), and copolymers, as their coating materials. The coatings are dissolved in fast-drying chlorinated hydrocarbon solvents and are applied to a variety of substrates including urea, diammonium phosphate [7783-28-0], potassium sulfate [7778-80-5], potassium chloride [7447-40-7], and ammonium nitrate / potassium sulfate-based N–P–K fertilizers.

Chisso-Asahi produces a number of different fertilizer grades. Because the thermoplastic polymers used are highly impermeable to water, release controlling agents such as ethylene–vinyl acetate and surfactants are added to the coating to obtain the desired diffusion characteristics. Coating thicknesses are essentially the same for all products with the release pattern being controlled by the level of release controlling agent. Release rates can also be altered by blending talc resin into the coating. Release patterns are characterized by a water leach test conducted at 25°C. A designation T-180 means the product releases 80% of its nutrient over 180 days when placed in water at room temperature. Products ranging from T-90 to T-360 are marketed.

As with other polymer-coated fertilizers, nutrients are released by diffusion through the coating. The various releasing agents incorporated into the coating change the permeability characteristics. The amount of release agent contained in the coating determines how fast the nutrients will diffuse. Typical of most polymer-coated fertilizers, the release is largely affected by temperature. However, the manufacturer attempts to minimize this effect by dispersing mineral fillers into the coating. Figure 6 shows the turfgrass growth response to a typical Asahi coated fertilizer marketed as Nutricote in the United States.

Chisso-Asahi uses a spouted bed process for the production of their coated materials (12). A 12,000 t / yr facility is located in Japan. The semicontinuous process consists of two batch fluid-bed coaters. A dilute polymer solution is prepared by dissolving 5% polymer and release controlling agent into a chlorinated hydrocarbon solvent such as trichloroethylene. The solution is metered into the spouted bed where it is applied to the fertilizer core. Hot air, used to fluidize the granules, evaporates the solvent which is recovered and reintroduced into the process. Mineral talc, when used, is either slurried into the polymer solution or introduced into the fluidizing air.

Asahi Chemical Industries uses this process to coat N–P–K fertilizers, marketed as Long, High Control, and Nutricote.

Urea and ammonium sulfate [7783-20-2] are coated by Chisso Co. under the trade names LP Cote and Meister. All U.S. consumption of these products is sourced from Japan. Chisso-Asahi products are marketed through very specific distribution channels (Table 5). Coated N–P–K products are marketed primarily to commercial nurseries and greenhouses. Coated urea products are marketed in blends to commercial nurseries, as well as to professional turf and strawberry growers.

Table 5. Distributors of Chisso-Asahi PCF

Distributor	Product type	Trade name
Plantco, Inc.	coated N–P–K	Nutricote
O.M. Scott & Sons	coated N / K	ProKote ^a
Vigoro Industries	coated urea	Escote
		Woodace ^b
Helena Chemical	coated urea	Meister

^a Blended with uncoated nutrients.

^b Also contains IBDU.

Exxon Chemical Canada. Exxon Research and Engineering Co. has developed a coating technology based on neutralized ionic elastomers, ie, the neutralization of various ionomers with metal cations such as zinc. PCT application (PCT / US87 / 00979) describes zinc neutralized sulfonated EPDM elastomers (see ELASTOMERS, SYNTHETIC—ETHYLENE PROPYLENE DIENE RUBBER) which have moisture barrier properties three orders of magnitude better than polyurethane. The barrier properties can be further improved by blending with other elastomers, such as styrene-4-vinylpyridine (13), or they can be made biodegradable by neutralizing with an amine-terminated caprolactone, instead of zinc (14,15). A second group of zinc neutralized ionic elastomers, ie, zinc-ethylene–methacrylic acid, has been described (16). Exxon Chemical Canada has developed a line of controlled release fertilizers, currently in the

development stage, which use the Exxon technology.

Exxon Chemical Canada's products are distinguished from many of the other polymer-coated fertilizer technologies by very thin coatings. Using ultrathin coatings of 2 to 4%, Exxon can control the release of fertilizers over 2- to 12-month time periods. The coating technology is adaptable to a number of fertilizer materials, with urea being the primary substrate used. Product release patterns are altered by the thickness of the coating as well as the composition of the polymer.

Exxon products appear to release via a unique mechanism. Like other polymer-coated technologies, the penetration of water into the granule is purely by diffusion. However, as water enters the particle, an osmotic pressure is created as the fertilizer is solubilized. This pressure causes an expansion of the elastomeric coating and the particle swells to many times its original diameter. As the particle swells, the coating becomes increasingly thinner to the point where it cannot contain the internal pressure and the nutrient is released.

Agronomically, Exxon products have been successful on a number of crop applications under varied climatic conditions from the plains of Canada to the heat and humidity of Florida.

Exxon currently operates a small developmental unit located in Redwater, Alberta, Canada. It utilizes a sophisticated polymer coating of neutralized ionic elastomers. The polymeric materials are dissolved in a fast-drying organic solvent, such as toluene, prior to application to the fertilizer. On the small-scale developmental unit, the polymer solution is applied to the fertilizer substrate in a batch fluid bed. The fluidizing air strips off the organic solvent leaving a thin but tough elastomeric film around each particle. The organic solvent is reclaimed and recycled back to the process for subsequent use.

Exxon Chemical Canada is currently test marketing this developmental technology on a number of applications including high value vegetable crop, turf, and ornamental nurseries.

O.M. Scott & Sons. The O.M. Scott & Sons Co. (Scotts) has developed a series of coated products which utilize copolymer blends of vinylidene chloride copolymerized with methyl methacrylates, acrylonitriles, methyl acrylates, and or vinylidene–vinyl chloride monomers.

These coated products contain from 5 to 15% coatings, are free-flowing, lack tackiness, and are resilient to handling. The nutrient release mechanism is a combination of diffusion and flow induced crystallization. Water vapor first must diffuse through the polymer coating to solubilize the fertilizer core. The osmotic pressure created by the solubilized core is thought to cause a change in the morphology of the polymer coating, from amorphous to crystalline. The crystallinity creates microchannels through which the fertilizer solution can be transported.

Scotts technology (17) uses fluid-bed (Wurster column) technology to apply polymeric coatings to a number of fertilizer substrates including urea, potassium nitrate, potassium sulfate, and monoammonium phosphate (MAP). The coating material is applied as a water-borne latex onto the fluidized substrate. As the substrate is fluidized with warm air (40–50°C), water is driven off and the latex coalesces into a continuous film around the fertilizer particle. The particular latex compositions used have selected glass-transition and blocking temperatures, which enable quick removal of the water before the soluble fertilizer core dissolves. This obviates the need to use precoats prior to the latex application.

Scott latex-coated products are designed for high temperature environments and long cycle crops such as commercial ornamental production and specialty agriculture.

Pursell Industries. In 1988, Pursell Industries acquired the production and marketing rights to a new developmental coating technology from Melamine Chemicals, Inc. This technology, known as the reactive layer coating (RLC) process, polymerizes two reactive monomers as they are applied to the fertilizer substrate in a continuous coating drum. These *in situ* reactive layer polymerizations form an ultrathin membrane coating which controls nutrient release by osmotic diffusion.

A number of products are being marketed under the trade name POLYON. These include coated basic fertilizer materials, ie, urea, potassium nitrate, potassium sulfate, potassium chloride, ammonium sulfate, ammonium phosphate, and iron sulfate, in various particle sizes. Coatings weights on urea vary from 1.5 to 15%, depending on the release duration desired. Table 6 lists typical products.

Table 6. Polyon Products

POLYON-coated basic material	Coating, wt %	Analysis grade ^a
urea	4	44-0-0
urea	8	42-0-0
microurea	11	41-0-0
potassium nitrate	5	12-0-42
potassium chloride	3	0-0-60
ammonium sulfate	5	20-0-0

^a Analysis grade refers to the primary nutrient content of a fertilizer. Nitrogen content is expressed as % elemental N. Phosphorus content is expressed as % P₂O₅, and potassium as % K₂O, given in the order N–P–K.

Nutrients are released from POLYON-coated fertilizers by osmotic diffusion. The RLC process permits application of ultrathin, hence lower cost, membrane coatings which distinguishes this technology from many other polymer-coated fertilizers. The coating thickness determines the diffusion rate and the duration of release. POLYON-coated urea at a 4% coating (44% N) will release at twice the rate and will have half the duration as an 8% coating (42% N).

During the manufacturing process for POLYON (18,19), the fertilizer substrate is coated with an excess of liquid diphenylmethane diisocyanate. This excess allows the diisocyanate to react with both the substrate surface and a layer of any one of several liquid polyester, or other polyols, which are subsequently applied over it. The two solvent-free liquids react, polymerizing *in situ*, to form a polyurethane coating which also is chemically bonded to the fertilizer core. The advantage of RLC technique is that the liquid coating materials can be applied and polymerized to the desired coating thickness in a continuous coating process drum without the need for solvents and associated recovery equipment. As a result, production costs are lower than many other commercial polymer-coated fertilizer technologies.

POLYON-coated products are marketed for turf and commercial nursery applications and specialty, medium to high cost, agricultural crop applications.

Haifa Chemical. Haifa Chemicals, Ltd. (Haifa, Israel) has developed a line of resin-coated products. Fertilizer granules are heated in a rotating pan and treated with a fatty acid and metal hydroxide, such as stearic acid [57-11-4] and calcium hydroxide [1305-62-0]. The two react to form a coating of the metal salt of a fatty acid, such as calcium stearate [1592-23-0]. Multiple layers of the fatty acid salt are reacted *in situ* followed by the application of a paraffin topcoat. Coating weights are relatively large compared to other technologies, but this is offset by the comparatively low cost of the coating materials. Substrates coated include potassium nitrate, urea, and triple superphosphate (TSP). The various coated components are blended together into different grades which are being marketed in the United States, under the name Multicote, for a variety of applications including specialty agriculture, turf, and commercial ornamental production (Table 7). The agronomic response to Haifa-coated urea (multicote 4 (29-0-0)) on turf is compared to other polymer-coated fertilizers in Figure 6.

Table 7. Haifa Products Marketed as Multicote

Product	N, %	P ₂ O ₅ , %	K ₂ O, %
Multicote 4(1-1-1)	12	12	12

Multicote 4(3-1-1)	20	6	6
Multicote 4(2-1-1)	14	7	14
Multicote 4(1-0-1)	18	0	18
Multicote 4(3-1-2)	17	6	12
Multicote 4(KNO ₃)	9	0	32
Multicote 4(urea)	29	0	0

Other Polymer-Coated Technologies. A number of other polymer-coated fertilizer technologies have been developed and are being marketed in Japan (Table 8). These products are used primarily in Japan on paddy rice and upland field agriculture.

Table 8. Japanese Polymer-Coated Fertilizer Technologies

Company	Coating	Type	Substrate	Trade name
Showa Denko	alkyd resin	thermoset	N–P–K	Sho-Cote
Central Glass	alkyd resin	thermoset	urea, N–P–K	Cera-cote
Nissan Chemical		thermoplastic	N–P–K	
Sumitomo Chemical		thermoplastic	urea	
Kyowa Hakko		thermoplastic	N–P–K	

Polymer/Sulfur-Coated Fertilizers. Polymer sulfur coated fertilizers (PSCF) are hybrid products that utilize a primary coating of sulfur and a secondary polymer coat. These fertilizers were developed to deliver control release performance approaching polymer-coated fertilizers, but at a much reduced cost. Sulfur is employed as the primary coating because of its low cost. Low levels of a polymer surcoat are used to control nutrient release rate. Unlike the soft wax (hydrocarbon oil) sealants used to cover imperfections in the sulfur coatings of SCUs, the polymers in this class of products are chosen to provide a continuous membrane through which water and nutrients must diffuse. The water permeability characteristics of the polymer controls the rate of water diffusion into the particle. The combination of the two coatings permits a positive cost benefit value over products with singular coatings of sulfur or polymer.

Products are granular, free-flowing, and dust-free by nature, since no flow conditioner dust is used as with sulfur-coated fertilizers. They possess excellent abrasion resistance and handling integrity. Since the outer coating is a hard polymer, the products do not leave waxy residues on material handling and application equipment.

The nutrient release mechanism is through a combination of diffusion and capillary actions. Water vapor must first diffuse through the continuous polymeric membrane layer. The rate of diffusion is controlled by the composition and thickness of the polymeric film. Once at the sulfur polymer interface, the water subsequently penetrates the defects in the sulfur coat through capillary action and solubilizes the fertilizer core. The solubilized fertilizer then exits the particle in reverse sequence. This diffusion controlled mechanism permits greater uniformity in nutrient release as compared to the typical matrix release of sulfur-coated fertilizers. The agronomic advantages of this are reduced surge growth after application and longer residual of up to six months. The difference in nutrient release between PSCF, eg, Scott's Poly-S, and SCU is illustrated by the comparative growth response of grass in Figure 7. In addition, the combination coating renders the nutrient release much less temperature sensitive than most polymer-coated fertilizers.

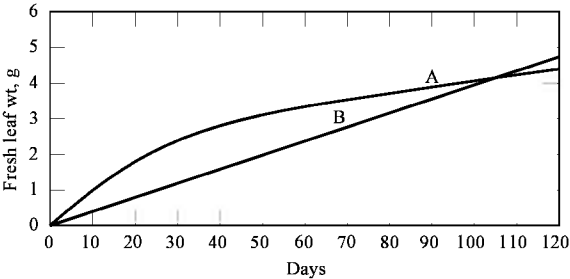


Fig. 7. Response of Kentucky bluegrass to polymer sulfur-coated fertilizers (PSCF) and sulfur-coated urea (SCF). Application of 2 lbs N /1000 ft² (9.76 g /m²). A is PSCF (Scott POLY-S); B, SCU.

Courtesy of The O.M. Scott & Sons Co.

PSCFs are manufactured by a two-step coating process. The first step applies the sulfur coating by techniques described in the sulfur-coated fertilizer section; the sealant coat and conditioner are not applied. The application of the polymer is accomplished by various methods depending on the nature of the polymer applied. Melt and reaction-coated polymers are typically applied in coating drum operations, whereas solvent borne, including latex, polymers are applied in a fluid-bed operation which utilizes heated air to strip off the solvent. The O.M. Scott & Sons Co. has commercialized a line of products under the trade name Poly-S, and Pursell Industries is marketing a PSCF under the trade name SulfurKote II.

PSCFs serve a number of markets including turf and garden, specialty agriculture, and commercial ornamental production.

BIBLIOGRAPHY

1. H. Kadowacki, *Bull. Chem. Soc. (Jpn.)* **11**, 248–261 (1936).

2. K. G. Clark, J. Y. Yee, and K. S. Love. *Indust. Eng. Chem.* **40**, 1178–1183 (July 1948).

3. AOAC 945.01, 15th ed., 1990.

4. AOAC 983.01, 15th ed., 1990.

5. U.S. Pat. 3,677,736 (July 18, 1972), R. Formaini (to Allied Chemical Corp.).

6. U.S. Pat. 4,089,899 (May 16, 1978), D. S. Greidinger (to Fertilizers and Chemicals, Ltd.).

7. U.S. Pat. 3,989,470 (Nov. 2, 1976), R. Czurak and R. M. Thompson (to The O. M. Scott and Sons Co.).

8. U.S. Pat. 5,102,440 (Apr. 7, 1992), W. E. Gallant, H. M. Goertz, and R. Martinez (to The O. M. Scott and Sons Co.).

9. AOAC 970.04, 15th ed., 1990.

10. *New Developments in Fertilizer Technology*, 14th Demonstration, TVA: National Fertilizer Development Center, Muscle Shoals, Ala., Oct. 5–6, 1983, pp. 30–33.

11. U.S. Pat. 3,903,333 (Sept. 2, 1975), A. R. Shirley (to Tennessee Valley Authority).

12. U.S. Pat. 4,019,890 (Apr. 27, 1977), T. Fujita and co-workers (to Chisso Asahi Fertilizer Co., Ltd.).

13. U.S. Pat. 4,701,204 (Oct. 20, 1987), L. I. Duvdevani, P. V. Manalastas, E. M. Drake, and W. A. Thaler (to Exxon Research & Engineering Co.).

14. U.S. Pat. 4,741,956 (May 3, 1988), W. A. Thaler, P. V. Manalastas, E. M. Drake, and R. D. Lundberg (to Exxon Research & Engineering Co.).
15. U.S. Pat. 4,788,081 (Nov. 29, 1988) (to Exxon Research & Engineering Co.).
16. U.S. Pat. 4,801,498 (Jan. 31, 1989) (to Exxon Research & Engineering Co.).
17. U.S. Pat. 5,089,041 (Feb. 18, 1992), H. E. Thompson and R. Kelch (to the O. M. Scott & Sons Co.).
18. U.S. Pat. 4,711,659 (Dec. 9, 1987), W. P. Moore.
19. U.S. Pat. 4,969,947 (Nov. 13, 1990) (to Melamine Chemicals, Inc.).

General References

Farm Chemicals Handbook, Meister Publishing, Willoughby, Ohio, 1990.
U.S. Pat. 4,969,947 (Nov. 13, 1990).
New Fertilizer Materials, Centre International des Engrais Chimique (CIEC), Noyes Development Corp., Park Ridge, N.J., 1968.
"Official Methods of Analysis", *J. Assoc. Off. Anal. Chem.*, (15) (1990).
Official Publication of Association of American Plant Food Control Officials (AAPFCO), Association of American Plant Food Control Officials, West Lafayette, Inc., No. 44, 1991.
Proceedings of the Symposium of Fertilizer, Present and Future, Japanese Society of Soil Science and Plant Nutrition, Sept. 25–26, 1989.
A. D. Davidson, *J. Assoc. Off. Anal. Chem.*, 769–774 (1983).
U.S. Pat. 3,223,518 (Dec. 14, 1965), L. I. Hanson (to Archer-Daniels-Midland).
J. T. Hays, "Controlled Release Nitrogen Fertilizers," *Manual Fert. Proc.* (1987).
S. P. Landels, M. M. Smart, J. Bakker, and J. Shimamoto, "Controlled Release Fertilizers and Nitrification Inhibitors," *Chemical Economics Handbook*, SRI International, Menlo Park, Calif., 1989.
S. P. Landels, *Proceedings of the Controlled Release Fertilizer Workshop*, TVA: National Fertilizer and Environmental Research Center, July 30–31, 1991, pp. 87–101.
P. V. Manalastas, W. A. Thaler, E. N. Drake, and I. S. Pasternak, PCT Publication W087 06575, Exxon Research & Engineering Co., Nov. 5, 1987.
Eur. Pat. Appl. 0,276,179 (July 1, 1988), G. Pipko (to Haifa Chemicals, Ltd.).
S. Shoji, and A. Gandizer, *Controlled Release Fertilizers with Polyolefin Resin Coating*, Konno Printing, Sendai, Japan, 1992.

Harvey M. Goertz
The O.M. Scott and Sons Company

PHARMACEUTICAL

Controlled-release dosage forms enhance the safety, efficacy, and reliability of drug therapy. They regulate the drug release rate to control drug action and reduce the frequency of drug administration to encourage patients to comply with dosing instructions. Conventional dosage forms often lead to wide swings in serum-drug concentrations. Most of the drug content is released soon after administration, causing drug levels in the body to rise rapidly, peak, and then decline sharply. For drugs whose actions correlate with their serum-drug concentration, these sharp fluctuations often cause unacceptable side effects at the peaks, followed by inadequate therapy at the troughs (Fig. 1). Administering smaller doses at more frequent intervals can damp concentration fluctuations, preventing over- and underdosing, but the inconvenience may cause patients to skip or delay doses, degrading efficacy. Compliance, measured as the percentage of patients taking 95% to 105% of prescribed oral medications, decreased from 67% with once-a-day regimens to 22% with four-times-a-day regimens (1). These difficulties are especially pronounced with drugs that have short half-lives and a narrow range of safe and effective concentrations. Precisely controlled release is particularly valuable for such agents (see DRUG DELIVERY SYSTEMS).

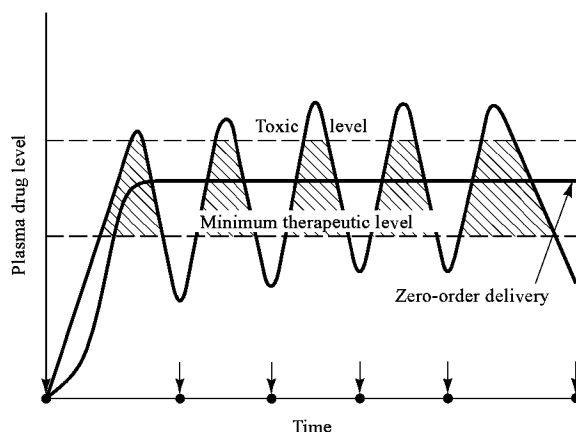


Fig. 1. Zero-order (controlled) delivery versus first-order (immediate-release) delivery (repeated administration). In zero-order delivery, the drug is released at a constant rate within the therapeutic range. In first-order delivery, each administration of the drug (represented by $\cdot$) causes serum-drug concentrations to rise to a peak and then decline sharply, resulting in over- and underdosing.

Courtesy of ALZA Corp., Palo Alto, Calif.

The term controlled release technology generally refers to a variety of methods used to exert various degrees of control over drug release. The U.S. Food and Drug Administration (FDA) defines controlled release dosage forms as those formulations designed to release active ingredient(s) at rates that differ significantly from their corresponding immediate release forms (2). The indexing system of *Index Medicus* supports this broad definition by classifying these dosage forms as delayed-action preparations, and *Excerpta Medica* uses the index term sustained release preparations. The pharmaceutical literature contains a profusion of terms, including delayed-action, extended action, gradual release, prolonged release, protracted release, slow release, sustained release, depot, retard, and timed-release dosage forms, all emphasizing an increased duration of action (3).

To avoid confusion, several researchers have incorporated therapeutic intention into the definition of controlled release (4–7). Thus, controlled-release pharmaceuticals release drugs *in vivo* according to a predictable, therapeutically rational, programmed rate to achieve the optimal drug concentration in the minimal time (4). Specification by release rate complements specification by quantity; jointly considered, they fix the duration of drug release. Therefore, the drug's duration of action can become a design property of a controlled release dosage form rather than an inherent pharmacokinetic property of the drug molecule.

Under this more precise definition, controlled release drug delivery not only provides a predictable, patterned action, but controls the rate of drug release for a predetermined period. Controlled release systems deliver drugs at a constant rate (zero-order release), a predictably constant declining rate (first-order release), or some other specified rate or pattern of rates to achieve the optimal serum-drug concentrations. In many clinical situations, a constant serum-drug level is desirable and zero-order drug delivery systems are appropriate. However, the body responds to certain conditions, eg, heart rhythm disorders and hormone deficiencies, and to certain substances according to strong circadian or other nonconstant patterns (8,9); in these cases, patterned or self-regulated delivery can enhance therapeutic efficacy. For a review of the pharmacokinetic and pharmacodynamic basis of controlled drug delivery, see Reference 10.

Most controlled release dosage forms administer drug according to their design, whether conceptually simple, eg, a fixed release rate for a fixed amount of time, or complex, eg, several different rates for different amounts of time. Alternatively, closed-loop systems contain a sensor to monitor drug concentration or to administer drug according to a biological need (11).

Although the field of controlled release technology is only a few decades old, the late 1980s and early 1990s have seen an explosion in research aimed at creating new drug delivery systems (12) as well as numerous publications that discuss controlled release. In addition, journals and societies have been established that are devoted to the advancement of drug delivery systems.

A principal focus of research and development in controlled release technology is the delivery of proteins and peptides (13,14), owing in part to biotechnological advances that allow the large-scale production of therapeutically important natural proteins, such as insulin [9004-10-8], and analogues. Because these substances tend to be broken down and inactivated in the gastrointestinal tract before they can be absorbed in therapeutic amounts, researchers are looking into enhancing oral delivery, eg, using permeation enhancers to increase absorption, as well as using less conventional delivery sites and routes of administration including the rectum, skin, oral cavity, nose, eyes, lungs, uterus, and vagina.

Delivery Sites, Portals, and Systems

Controlled release pharmaceuticals provide local or systemic treatment and differ in design depending on their use and route of administration. The potential sites for controlled delivery of therapeutic agents and some of the systems designed for those sites merit discussion.

Gastrointestinal Tract. The gastrointestinal (GI) tract is the most familiar and widely used portal for drug delivery, largely because of the ease with which medications can be administered.

Stomach, Small Intestine, and Colon. These organs provide accessible portals for local delivery of therapeutic agents, including laxatives,

antidiarrheal medications, gastroprotective agents such as sucralfate [54182-58-0], and anthelmintics.

They also are important portals for systemic therapy. However, many variables can influence drug dissolution and absorption in these areas, including rate of gastric emptying, intestinal motility, mass and pH of intestinal contents, and condition of the absorbing surfaces (15–17). These variables, in turn, can be affected by the patient's disease, posture, and eating habits, and even by such aspects of the treatment as the timing of doses (11).

Dosage forms have been designed to optimize the localized actions or systemic absorption of drugs by targeting specific GI tract areas for drug release. Many dosage forms have pH-sensitive enteric coatings that prevent digestion in the stomach, reducing local irritation and ensuring that the drug core does not dissolve until it reaches the small intestine or colon. Numerous pH-sensitive, polymer-coated oral preparations deliver 5-aminosalicylic acid [89-57-6] to the colon to treat chronic inflammatory bowel disease. Marketed controlled-release preparations of 5-aminosalicylic acid include Asacol (available in the United Kingdom and Sweden) and Claversal (available in Denmark), coated with Eudragit S [117385-53-2] and Eudragit L [76688-80-7], respectively, and Pentasa (available in Denmark), which is coated with ethylcellulose [9004-57-3] (18,19). A different approach to colon-targeted delivery is the use of Eudragit RS [121381-94-0] microspheres loaded with 5-aminosalicylic acid and incorporated within a colon-targeted device (20). Another colon-targeted system in research is a miniature osmotic pump, the Osmet (ALZA Corp.) delivery system, coated with a substance insoluble in low pH gastric fluids but soluble in the higher pH of the colon. The system delivers the drug approximately 2 to 4 h after passing through the stomach (21). A colon-targeted dosage form with multiple small osmotic systems loaded in a gelatin capsule delivers beclomethasone dipropionate [5534-09-8]. *In vitro*, the system incorporates a 12-h delay before delivering the drug for 3 h at an increasing rate and then for 20 h at a constant rate (22).

Permeation enhancers are used to improve absorption through the gastric mucosa. For example, oral delivery of insulin (mol wt 6000) has been reported from a water-in-oil- emulsion containing lecithin, nonesterified fatty acids, cholesterol [57-88-5], and the protease inhibitor aprotinin [9087-70-1] (23).

The design and fabrication of numerous commercially available controlled-release oral dosage forms have been widely reviewed (17,24–27).

Rectum. Local treatment of conditions and diseases of the rectum, including hemorrhoids and local inflammation, usually involves drug-containing suppositories or enemas.

As a portal for systemic delivery, the rectum has advantages over other sites in the GI tract; for example, bypassing the upper gastrointestinal tract and the liver avoids enzymatic degradation and first-pass metabolism, in which some of the drug is lost during its first pass through the GI tract and the liver. Drugs delivered rectally include steroids, sedatives, antihistamines, tranquilizers, antiemetics, and headache medications. Poly(vinyl alcohol) hydrogels for rectal administration have been designed to increase the bioavailability of propranolol [525-66-6] (28). Numerous drugs, including antipyrine [60-80-0], theophylline [58-55-9], propranolol, and nifedipine [21829-25-4], have been administered via the rectum with the Osmet module (29). An oral controlled-release morphine [57-27-2] tablet (MSContin) has been enclosed in a gelatin capsule and given as a rectal suppository to terminally ill patients who cannot tolerate oral medications (30). Controlled-release morphine in a hydrogel has been tested in normal volunteers (31), and slow-release indomethacin [53-86-1] suppositories have been given to relieve postoperative pain after orthopedic surgery (32).

Skin. The skin's unique molecular transport and barrier properties pose a challenge for transdermal drug delivery. Diffusion of drugs through the stratum corneum, the outer layer primarily responsible for the skin's limited permeability, varies by drug, by skin site, and among individuals. Until recently, virtually all drugs applied to skin were topical treatments.

Now, however, dosage forms that use the skin as a reliable portal for systemic drug delivery are available. Despite skin's limited permeability, the transdermal route is appealing because it avoids first-pass metabolism and minimizes the dose necessary to achieve therapeutic serum-drug levels. To be suitable for systemic therapy from a transdermal delivery system, a drug must be nonirritating and potent and have a short half-life, low molecular weight, and low melting point (33). Creams and ointments that provide systemic drug delivery, such as estradiol [50-28-2] and progesterone [57-83-0] (Oestrogel and Progestogel) for menopausal symptoms, are messy and difficult to administer in consistent amounts, and have unpredictable absorption rates.

Passive transdermal delivery systems on the market tend to be either matrix or membrane controlled. In matrix devices, the structural and molecular characteristics of the drug-polymer matrix determine drug release. Examples of polymer matrix-controlled diffusional systems for angina prophylaxis include Nitro-Dur and Nitrodisc, which provide transdermal delivery of nitroglycerin [55-63-0], and Frandol, a tape that releases isosorbide dinitrate [87-33-2]. Matrix diffusional systems have been used for delivering drugs with a wide therapeutic index.

Membrane-controlled diffusional systems enhance transdermal delivery by increasing control over serum–drug concentrations. They are particularly suitable for delivering compounds with a narrow therapeutic index. Because it is predominantly the membrane rather than the skin that governs drug release, the system reduces the problem of differing skin absorption rates. Commercially available membrane-controlled systems are typically administered from once a day to once a week and include Transderm-Nitro (nitroglycerin) for angina, Catapres-TTS (clonidine [4205-90-7]) for hypertension, Estraderm (estradiol) for menopausal symptoms, Transderm Scōp (scopolamine [51-34-3]) for motion sickness, and Duragesic (fentanyl [437-38-7]) for severe chronic pain requiring narcotics. These were introduced by ALZA Corp.

Various strategies are available to overcome the barrier properties of skin. Permeation enhancers can facilitate transdermal transport of drugs with low percutaneous flux. Estraderm, for example, delivers 0.3 mL of the permeation enhancer ethanol [64-17-5] for every 4 mg estradiol. Other permeation enhancers, including Azone [59227-89-3] and esters of saturated and unsaturated fatty acids, are in experimental use (34). Numerous potential permeation enhancers have been reported in the literature, including various alcohols, propylene glycol [57-55-6], and sodium lauryl sulfate [151-21-3] (33). Ultrasound (phonophoresis), electrical charge (electrotransport), and bioconvertible drug precursors, or pro-drugs, also are being investigated as ways to enhance drug transport through the skin (28).

Oral Cavity. The oral mucous membranes are subject to many diseases and chronic lesions, for which a variety of local treatments are available. Rinses and swabbed liquids are effective but must be applied frequently. Slow-dissolving lozenges, eg, Mycostatin, which delivers the antifungal antibiotic nystatin [1400-61-9], and chewing gum that delivers the antifungal miconazole [22916-47-8] (35), can sustain drug release in the oral cavity, reducing the need for frequent reapplication (36). Incorporation of drugs in inert polymerized acrylics prolongs drug delivery for several days. The Actisite (ALZA Corp.) periodontal fiber, with an ethylene–vinyl acetate copolymer, is designed for placement in the periodontal pocket, where it releases tetracycline hydrochloride [64-75-5] at a controlled rate for 10 days for treatment of periodontal disease. This system is awaiting market approval from the U.S. Food and Drug Administration. Intrapocket devices in development include cellulose-based hollow fibers, poly(ethyl methacrylate) [9003-42-3] strips, and ethylcellulose [9004-57-3] slabs that release chlorhexidine gluconate [18472-51-0], metronidazole [443-48-1], minocycline [10118-90-8], and tetracycline [60-54-8]. Biodegradable intrapocket devices are also being tested clinically, eg, a cross-linked protein matrix delivers chlorhexidine [55-56-1] for up to 90 h (37), and poly(hydroxybutyric acid) [52352-27-9] strips deliver tetracycline or metronidazole for approximately 4 days (38). Other degradable delivery devices are being developed, but no clinical data are available yet (30).

Oral mucosal membranes provide a port for systemic therapy as well. Nitroglycerin sublingual tablets (Nitrostat) abort acute angina attacks; methyl-testosterone [58-18-4] buccal tablets (Android 5) are indicated for testosterone [58-22-0] replacement therapy (39); and nicotine [54-11-5] gum (Nicorette) aids in smoking cessation.

Nose. Intranasal delivery of drugs, such as antihistamines and decongestants, to alleviate local symptoms is common. There is now increasing interest in using the nose as a portal for systemic drug delivery, particularly for the delivery of proteins and peptides (40). Several classes of nasal permeation enhancers have been investigated, including surfactants, such as polyoxyethylene-9-lauryl ether [977007-24-1], bile salts and bile salt derivatives, such as sodium taurodihydrofusidate [42907-93-7], and mucoadhesive systems (41). Controlled delivery from bioadhesive microspheres has been proposed to avoid the rapid clearance of therapeutic agents by nasal cilia. Such microspheres would have to be larger than 10 to 15 μm in order to remain in the nasal cavity and not travel to the lungs (42).

Eye. The eye is an easily accessible but problematic portal for local therapy, that is, for eye diseases or trauma, because constant tear flow and lacrimal–nasal drainage result in extensive drug loss. The most familiar local treatments (eye drops, suspensions, and ointments) require frequent application to achieve steady-state drug levels (43). Numerous approaches enhance local treatment by improving corneal permeation with ionophores, ion pairs, liposomes, or pro-drugs, or by reducing loss from drainage with viscosity-enhancing agents, suspensions, emulsions, ointments, or erodible or nonerodible matrices (44–48). The Ocuser (ALZA Corp.) ocular therapeutic system for the release of pilocarpine [92-13-7] to treat glaucoma was the first diffusion-controlled ocular system marketed (1975) (49). The Ocuser system is a small, flexible oval unit composed of a drug reservoir surrounded by a

biocompatible ethylene–vinyl acetate [24937-78-8] membrane (Fig. 2). The patient inserts the system under the eyelid, where it releases the drug for seven days. The Lacrisert system, a rod-shaped insert that the patient places under the eyelid once or twice daily, slowly delivers hydroxypropylcellulose [9004-64-2] to treat dry-eye syndrome. The Bio-Cor collagen shield is a therapeutic film composed of porcine scleral tissue (high in collagen) for treatment of eye infections. A physician uses forceps to place the film on the anesthetized eye, where it conforms to the corneal surface and slowly dissolves within 12, 24, or 78 h (44,50). The collagen shield has been used experimentally to deliver amphotericin B [1397-89-3] (51). Additional research activity includes development of an ocular polymeric drug delivery system for local therapy. The system utilizes the prepolymer α , ω -bis-(4-methacryloxybutyl)-poly(dimethylsiloxane) (M_2D_{30}) copolymerized with acrylate-based monomers for local therapy (52).

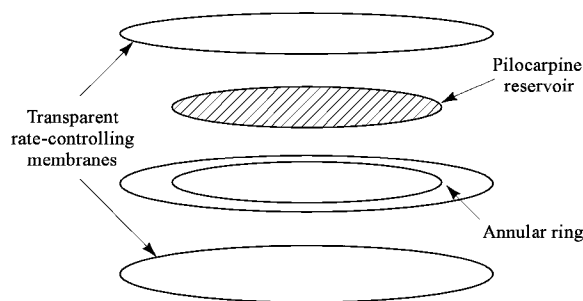


Fig. 2. Ocusert ocular therapeutic system. The Ocusert system releases pilocarpine at a controlled rate to treat glaucoma. In this schematic representation, the three disks and the ring are shown two-dimensionally. The patient inserts the Ocusert system under the eyelid, where it releases pilocarpine for seven days.

Courtesy of ALZA Corp.

No ocular products for systemic therapy are commercially available, but research is under way on ocular systems for the systemic delivery of therapeutic agents such as insulin (53).

Lungs. The lungs are easily accessible for both local and systemic treatment. Local treatment of various lung and bronchial diseases typically involves the inhalation of drugs such as sodium cromoglycate [15826-37-6], corticosteroids, pentamidine [100-33-4], and beta adrenoceptor agonists such as salbutamol [18559-94-9] and terbutaline [23031-25-6]. Solutions of these drugs are converted into aerosols and delivered with nebulizers or metered dose inhalers (54). Strategies to lengthen the duration of drug release into the respiratory tract include modifying a drug's molecular design, eg, salbutamol is modified to form salmeterol [89365-50-4]; changing the form of a drug, eg, from a salt suspension to a parent-free base; or incorporating drugs into liposomes (55,56).

The lungs' large, permeable surface makes systemic delivery possible. For example, an inhaler delivers 360 μ g per dose of aerosolized ergotamine tartrate [379-79-3] for migraine (54), and inhalant systems deliver anesthetic gases. Research is under way on the systemic delivery of proteins and peptides through the lungs (57,58).

Uterus. As a portal for drug delivery, the uterus has had fairly limited use. The only local treatment is through intrauterine devices (IUDs) of contraceptive metals or steroid hormones, which provide long-term contraception. The two IUDs available in the United States are a copper-containing IUD (Paragard T 380A) that can be used for six years, and an ethylene–vinyl acetate copolymer IUD from ALZA Corp. (Progestasert) that releases progesterone at a nearly constant rate of approximately 65 μ g per day for 1 year (see CONTRACEPTIVES). Two IUDs that release the progestin levonorgestrel [797-64-8], one for 5 years and one for eight years, are available outside the United States (59,60). No uterine dosage forms are available to deliver drugs for systemic therapy.

Vagina. The vagina is an accessible but little used route of drug administration. All commercially available vaginal delivery systems are for local therapy. Vaginal douches, suppositories, creams, foams, and ointments effectively deliver drugs for vaginal infections, contraception (spermicides), and vaginal dryness (estrogens). Controlled release prostaglandin E_2 [363-24-6] (Propress) for labor induction (cervical ripening) is available in the United Kingdom (61).

The permeability of the vaginal mucosa makes the vagina a route for systemic delivery as well. Vaginal rings that deliver contraceptive steroids for up to 21 days are being tested clinically (59,62), as are rings delivering estradiol for postmenopausal estrogen replacement (63).

Intramuscular, Intravenous, and Subcutaneous Portals. These routes require injections, infusion systems, or implants. Injections for local treatment are limited to a few common drugs, including procaine [59-46-1] or cortisone [53-06-5] to ease pain and stiffness in joints and bursas. For systemic therapy, a wide variety of drugs are available in intramuscular or subcutaneous injections for intermittent dosing. Sustained systemic treatment can, of course, be achieved by intra-arterial or intravenous infusions with catheters attached to containers or pumps. Several miniaturized infusion pumps that attach to the patient, allowing prolonged ambulatory care, are now available, including the Baxter Infusor, the Mill Hill Infuser, and the Auto-Syringe (64). Sustained systemic treatment is also available through implants such as the Norplant system, a 5-yr contraceptive recently approved for marketing in the United States. The Norplant system is a set of six flexible, closed capsules made of Silastic (dimethylsiloxane–methyl-vinylsiloxane copolymer), each 34 mm long and 2.4 mm in diameter and containing 36 mg of levonorgestrel. The capsules are inserted beneath the skin of the upper arm (59). For agents such as insulin, a vapor-pressure-powered implantable system about the size and shape of a hockey puck provides long-term continuous infusion (65,66).

Energy Sources for Controlled-Release

The design of a therapeutic system varies with its energy source as well as with its route of administration. For controlled drug delivery, energy sources are typically processes or properties, such as osmosis or elasticity, although some infusion pumps and transdermal diffusional systems also use electricity. The following section discusses primary energy sources for controlled-release technology and presents representative systems. Some overlap clearly exists; biodegradable devices, for example, often employ both biodegradation and diffusion processes for drug release.

Biodegradation. Since the 1970s, biodegradation, ie, the *in vivo* degradation of polymers to form biocompatible by-products (monomeric subunits or soluble polymer fragments) that are metabolized or excreted, has been investigated as an energy source for controlled drug delivery (67–72). Since the mid-1980s this field of research has expanded greatly, with researchers studying the release of numerous pharmaceutical agents from biodegradable polymers.

Three mechanisms for the biodegradation of polymers have been discussed (Fig. 3) (71). Mechanism I involves water-soluble polymers that are cross-linked by covalent bonds to make them insoluble. Hydrolytic cleavage of either the cross-links (type IA) or the backbone (type IB) yields water-soluble polymers or polymer fragments whose size depends on the density of the bonds being hydrolyzed. Mechanism II involves water-insoluble polymers that become soluble when pendent groups are hydrolyzed, ionized, or protonated. In Mechanism III, cleavage of the hydrolyzable bonds in the polymer backbone produces low molecular weight, water-soluble fragments. Actual biodegradation may be a combination of these mechanisms.

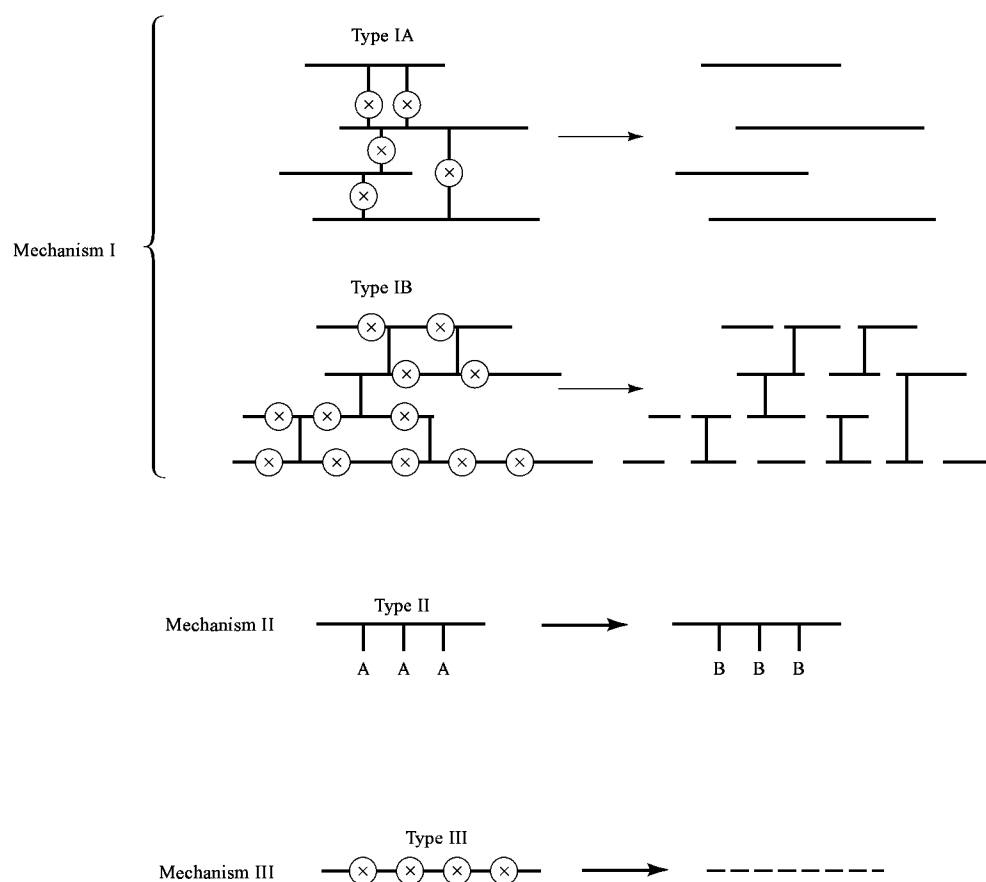


Fig. 3. Mechanisms for polymer degradation. The illustration is a schematic representation of three degradation mechanisms: I, cleavage of cross-links; II, hydrolysis, ionization, or protonation of pendent groups; III, backbone cleavage. Actual biodegradation may be a combination of these mechanisms.

Courtesy of CRC Press, Inc., Boca Raton, Fla.

Biodegradable polymers have several advantages as drug carriers. They allow controlled release over a designated period of time, can target a specific body site, cause little or no tissue reaction, and cause less discomfort and inconvenience than multiple injections. Implants of biodegradable polymers need not be removed. Finally they enable the use of drugs with short *in vivo* half-lives and may improve the bioavailability of drugs having low aqueous solubility.

In order to become useful drug delivery devices, biodegradable polymers must be formable into desired shapes of appropriate size, have adequate dimensional stability and appropriate strength-loss characteristics, be completely biodegradable, and be sterilizable (70). The polymers most often studied for biodegradable drug delivery applications are carboxylic acid derivatives such as polyamides; poly(α -hydroxy acids) such as poly(lactic acid) [26100-51-6] and poly(glycolic acid) [26124-68-5]; cross-linked polyesters; poly(orthoesters); polyanhydrides; and poly(alkyl 2-cyanoacrylates). The relative stability of hydrolytically labile linkages in these polymers (70) is as follows:



Homopolymers and copolymers of lactic acid [598-82-3] and glycolic acid [79-14-1], originally developed for use in absorbable sutures (qv) (for example, Dexon and Vicryl), are particularly appealing for controlled drug delivery because of their demonstrated safety. Many types of pharmaceutical agents, including narcotic antagonists, contraceptive steroids, antiinflammatory steroids, anticancer agents, antimalarials, antibiotics, peptides, and local anesthetics, have been incorporated into these polymers. Systems based on the polymers have been prepared in several forms, including cylinders and rods, particles and powders, microcapsules and microspheres, beads, films, fibers, and needles.

Drug release from biodegradable devices occurs by degradation of the polymer, by diffusion, or by a combination of these two processes. In bulk or homogeneous degradation, hydrolysis occurs throughout the polymer. In surface or heterogeneous degradation, hydrolysis is confined to the surface of the device. The mechanism and duration of drug release are in part determined by the method used to combine the active agents with the polymer. The manufacture of reservoir-type devices involves surrounding a core of the drug with a biodegradable membrane. Monolithic or matrix devices are fabricated by dispersing the drug homogeneously throughout the polymer. The release rate of a drug from a biodegradable matrix is affected by the amount of the drug present; its diffusibility and solubility in the polymer; the composition, molecular weight, and porosity of the polymer; the device geometry; and the presence of additives.

Biodegradable dosage forms are in various stages of clinical testing, and some are approved for marketing. Two products based on lactic and glycolic acid are available in the United States. Lupron Depot, marketed by TAP Pharmaceuticals, contains leuprolide acetate [74381-53-6], an analogue of LHRH (leuteinizing hormone releasing hormone [9034-40-6] (73); Zoladex, introduced by ICI Pharma in 1987 and now available in 27 countries (74), contains goserelin acetate, also an LHRH analogue (75). Both systems are injected once a month for the palliative treatment of prostate cancer. Another biodegradable system uses injectable microspheres or implants to deliver nafarelin [76932-56-4], a third LHRH analogue, formulated with erodible lactic acid–glycolic acid copolymers [34346-01-5] for approximately 1 month for the treatment of endometriosis or prostate cancer (76,77). Atrix Labs is developing a product containing a biodegradable polymer in liquid form that contains an antimicrobial or antibiotic agent combined with a solution of poly(glycolic acid) or a copolymer of lactic or poly(glycolic acid) (78). The liquid is injected into the periodontal pocket, where it solidifies, releasing agents such as tetracycline hydrochloride [64-75-5] or chlorhexidine diacetate [56-95-1] that aid in controlling the microorganisms that cause periodontitis. InSite Vision is developing a biodegradable system for treating ophthalmological conditions such as glaucoma or dry-eye syndrome. The liquid solidifies when introduced into the conjunctival sac underneath the lower eyelid (79).

Diffusion. Diffusional drug delivery systems utilize the physicochemical energy resulting from concentration differentials. Drug molecules diffuse through a polymer matrix or through a polymer membrane film from a region of high concentration to one of low concentration.

Matrix Diffusional Systems. In matrix diffusional systems, the drug is dispersed in an appropriate vehicle, usually a high viscosity lipophilic or hydrophilic polymer. The drug is generally in solid form, although it can be liquid. Such systems are inexpensive and easy to formulate, typically, by blending the drug into the polymer using conventional polymer mixing procedures and then shaping the adduct by extrusion or molding.

Release of drugs from the outer surface of matrix devices into the appropriate body site begins when dissolved drug molecules leave the surface of the system and move into the receptor. The developing concentration gradient causes more molecules to dissolve and diffuse into the system's surface. The undissolved drug particles replenish the concentration of dissolved drug molecules. In this process of dissolution–diffusion–release, the distance from the surface of the system to the undissolved drug continually grows, forcing the dissolving drug molecules to traverse an ever-increasing expanse.

Release kinetics of the matrix system have been derived (80) and refined (81):

$$\frac{dM_t}{dt} = \frac{A}{2} \frac{(2DC_sC_o)^{1/2}}{t^{1/2}} \tag{1}$$

where dM_t/dt is the rate of release of the drug from the surface of the system, A is the surface area of the system, D is the diffusion coefficient of drug molecules through the polymer, C_i is the solubility of the drug in the polymer, C_o is the total concentration (dissolved and dispersed) of the drug in the matrix, and t is time.

Matrix diffusional devices generally have C_o very much greater than C_i , so that at the end of the system's scheduled operating life some undissolved drug remains in the system. The release rate from such systems declines as a function of the inverse square root of time during most of the release interval; the plot of cumulative release versus $t^{1/2}$ is a straight line. Although this declining release rate is a limitation when constant drug delivery is required, it can be an advantage when the resultant plasma drug profile coincides with the body's natural rhythm for substances such as hormones (33). For example, after daily morning application of Testoderm (ALZA Corp.), a matrix transdermal system for treatment of hypogonadism awaiting FDA approval, resulting serum-drug levels mimic the circadian pattern of testosterone observed in normal young men (82).

Another matrix diffusional implant consists of an outer layer of micronized, crystalline 17 β -estradiol dispersed in silicone rubber over a nonmedicated, cylindrical silicone rubber core. The system, implanted subcutaneously in the ears of cattle, releases estradiol for up to 400 days with $t^{1/2}$ kinetics to improve growth rate and feed efficiency (83).

Although current matrix diffusional systems are most suitable for small-molecule compounds, it has been demonstrated (84) that solid hydrophobic polymers allow dispersed powdered macromolecules of nearly any size, for example, ethylene-vinyl acetate copolymers containing dispersed polypeptides, to be released for periods exceeding 100 days.

Membrane Diffusional Systems. Membrane diffusional systems are not as simple to formulate as matrix systems, but they offer much more precisely controlled and uniform drug release. In membrane-controlled drug delivery, the drug reservoir is intimately surrounded by a polymeric membrane that controls the drug release rate. Drug release is governed by the thermodynamic energy derived from the concentration gradient between the saturated drug solution in the system's reservoir and the lower concentration in the receptor. The drug moves toward the lower concentration at a nearly constant rate determined by the concentration gradient and diffusivity in the membrane (33).

The drug can be in dry powder form, dispersed in liquid, or in a solid polymer matrix. The membrane can be a solid, single-component polymeric film, a polymer blend, a microporous or macroporous film, or a film of hydrophilic polymer particles dispersed in a hydrophobic polymer matrix. The rate-controlling membrane and the active core may be combined by many different techniques, such as laminating the drug element and solid membrane as films, coating a shaped drug element with a volatile solution of the membrane polymer, form-fill-and-seal processing (used in producing single-serve condiment packages), microencapsulating the drug, filling a tubular membrane with a dissolved or suspended drug, or loading the drug into membrane capsules.

If the drug is enclosed within a polymer membrane, the release rate is represented by the equation

$$\frac{dM_t}{dt} = DK \frac{C_s}{l} \tag{2}$$

where K is the coefficient of the drug partition between the reservoir and the membrane, C_s is the solubility of the drug in the reservoir, l is the thickness of the membrane, and D is the diffusivity of the drug in the membrane.

A variety of membrane-controlled systems are available. Transdermal membrane-controlled systems, composed of a drug-impermeable backing, a drug reservoir, a rate-controlling polymeric membrane, an adhesive layer, and a protective peel strip, deliver numerous drugs, including scopolamine, nitroglycerin, estradiol, and fentanyl. Figure 4 shows a schematic illustration of a membrane-controlled transdermal system in place on the skin. Nicoderm (ALZA Corp.), the transdermal therapeutic system delivers controlled amounts of nicotine for 24 h as an aid in smoking cessation. Other systems include the Ocusert ocular system and the Progestasert intrauterine device.

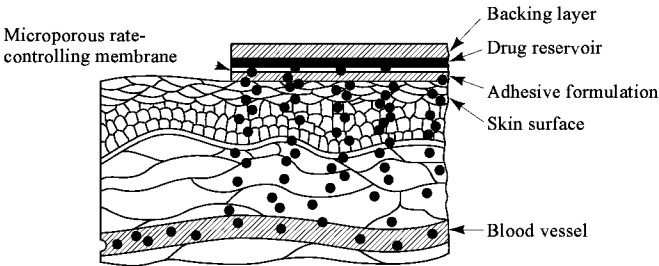


Fig. 4. Schematic of transdermal therapeutic system in operation. The drug diffuses through the intact skin into capillaries and is then carried into the general circulation.

Courtesy of ALZA Corp.

Elasticity. A portable infusion controlled-release device, the Baxter Infusor, utilizes energy stored in a distended, elastomeric rubber tube to deliver drug solutions at constant rates of flow through fixed resistances (85). The pharmacist inflates the drug reservoir with a loaded syringe inserted through a septum at the filling port, and the liquid drug is metered through a glass capillary. Figure 5 shows the tubular rubber reservoir fully inflated and ready to deliver a drug in liquid form. A description of the physicochemical properties of the elastomeric component and the design of the elastomeric reservoir is available (85–87). The system provides continuous delivery of cancer chemotherapeutic agents, for example 5-FU [51-21-8], interleukin-2 [85898-30-2], interferon [82115-62-6], methotrexate [59-05-2], cytosine arabinoside [147-94-4], doxorubicin [23214-92-8], bleomycin [11056-06-7], or cisplatin [15663-27-1] (see CHEMOTHERAPEUTICS, ANTICANCER). An alternative design allows patient-controlled delivery of analgesics such as meperidine [57-42-1] or morphine sulfate [64-31-3] (88) (see ANALGESICS, ANTIPYRETICS, AND ANTIINFLAMMATORY AGENTS).

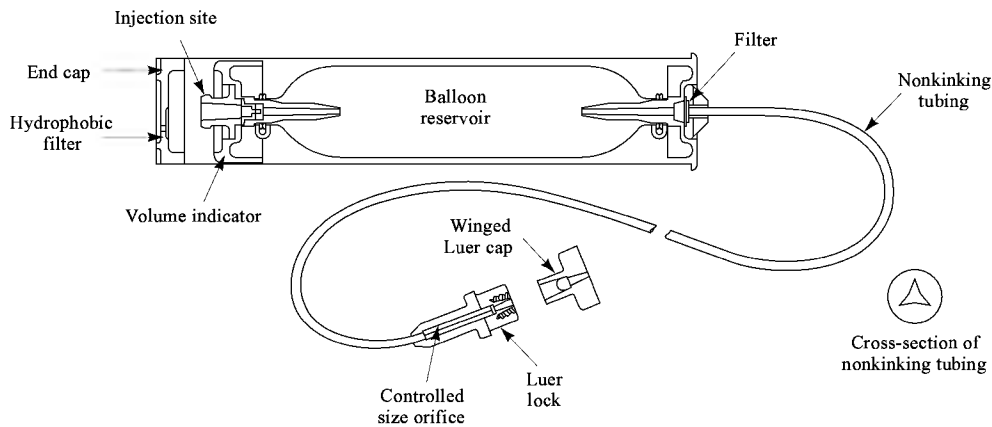


Fig. 5. Baxter Infusor Portable Infusion System. This portable, lightweight infusion system provides continuous, controlled drug delivery. The tubular rubber reservoir is shown fully inflated and ready to deliver the drug in liquid form.

Courtesy of Baxter Healthcare Corp., Round Lake, Ill.

Electrokinetics. Electrokinetics, the motion produced by an electrical field, is expected to play an increasingly important role in enhancing transdermal drug delivery. Commercial transdermal therapeutic systems, which typically rely solely on passive diffusion, cannot effectively deliver substances such as proteins and peptides that diffuse through the skin too slowly to achieve the desired dosing rate. In an electrical field, transdermal transport of ionized compounds occurs primarily through the sweat glands and hair follicles, the so-called shunt pathways through the stratum corneum, significantly increasing transport rates (89,90).

Electrically assisted transdermal drug delivery, ie, electrotransport or iontophoresis, involves the three key transport processes of passive diffusion, electromigration, and electroosmosis. In passive diffusion, which plays a relatively small role in the transport of ionic compounds, the permeation rate of a compound is determined by its diffusion coefficient and the concentration gradient. Electromigration is the transport of electrically charged ions in an electrical field, that is, the movement of anions and cations toward the anode and cathode, respectively. Electroosmosis is the volume flow of solvent through an electrically charged membrane or tissue in the presence of an applied electrical field. As the solvent moves, it carries dissolved solutes.

Mass transport of species (*i*) under an electrical field is governed by

$$J_i = -D_i \left(\frac{dC_i}{dx} \right) - \left(\frac{z_i D_i F C_i}{RT} \right) \left(\frac{d\phi}{dx} \right) + u C_i \tag{3}$$

where *J_i* is the flux of *i*, *D_i* is the diffusion coefficient, *C_i* is the concentration, *dC_i/dx* is the concentration gradient, *z_i* is the ionic charge, *F* is the Faraday constant, *dφ/dx* is the electrical field, *R* is the gas constant, *T* is the temperature, and *u* is the velocity of the convective flow via electroosmosis (91). Detailed discussions of electrotransport theory are available (92–95).

A simple electrotransport therapeutic system is composed of an electrical power source, an anode, and a cathode (Fig. 6). The anode and cathode are typically composed of multiple layers. For example, the anode can incorporate an electrode in conjunction with a skin-contacting hydrogel containing the salt of a cationic drug (*D*⁺). As the anode is discharged, the drug moves through the skin, carrying a fraction of the current. Simultaneously, anions move from the skin into the anode hydrogel. At the cathode the situation is reversed: anions move from the cathode hydrogel into the skin while cations move in the opposite direction.

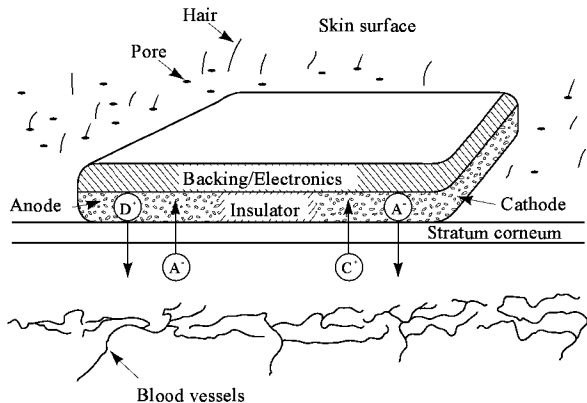


Fig. 6. Electrically assisted transdermal therapeutic system. Electrotransport facilitates passage of drugs (*D*⁺) through the skin and into adjacent tissue and the systemic circulation.

Courtesy of ALZA Corp.

Electrotransport technology offers a number of benefits for therapeutic applications, including systemic or local administration of a wide variety of therapeutic agents with the potential administration of peptides and proteins; long-term noninvasive administration, improving convenience and compliance; controlled release, providing a desired delivery profile over an extended period with rapid onset of efficacious plasma drug levels and in some cases reduced side effects; and a transport rate relatively independent of skin type or site. Additional benefits include easy inception and discontinuation of treatment, patterned and feedback-controlled delivery, and avoidance of first-pass hepatic metabolism.

Electrotransport of some 46 different compounds has been investigated (96). Perhaps most promising is the potential for transdermal delivery of peptides and proteins. Several researchers have reported preliminary results showing successful delivery of peptides and proteins including leuprolide [53714-56-0] (97), insulin (98,99), growth hormone releasing factor (100), and calcitonin [9007-12-9] (101).

Commercially available electrotransport systems are bulky and limited to acute applications (96). One example, the Drionic system used for the treatment of hyperhidrosis (excessive perspiration), is presoaked in water for 30 min before each 20- to 30-min treatment. Another system, the Phoresor, approved for the delivery of lidocaine [137-58-6] for local anesthesia, and of dexamethasone [50-02-2] for treatment of local inflammation such as bursitis or tendinitis, is powered by a 9 V replaceable battery and features a disposable, fillable drug electrode.

Future electrotransport therapeutic systems will differ substantially from those just described. They will draw on advances in microelectronics and transdermal system technology to provide transdermal therapy for compounds with low passive permeation rates, patterned or pulsed drug delivery,

on-demand drug administration such as patient-controlled analgesia, or closed-loop drug delivery using a biosensor. Electrotransport systems may be completely integrated and will likely be the size of a conventional transdermal system, that is, smaller than 50 cm², and easy to wear for extended periods.

Osmosis. Osmosis, a natural process in which molecules of a solvent move through a semipermeable membrane from a region of low to high solute concentrations, is the energy source for several commercially available therapeutic systems. Osmotic systems for human therapy have a solid core, usually shaped like a standard tablet, that contains the drug and often the osmotic agents. This core is coated with a semipermeable, rate-controlling membrane containing one or more laser-drilled orifices. In an aqueous environment such as the gastrointestinal tract, the osmotic activity of the core components establishes an osmotic activity gradient across the membrane, drawing water into the system at a rate controlled by the membrane's composition, thickness, and area. Drug delivery begins when the water enters the system to dissolve or suspend the drug; the drug solution or suspension flows out of the delivery orifice or orifices at a rate equal to the rate of water inflow through the membrane.

The simplest osmotic dosage form, ALZA Corporation's OROS elementary osmotic pump (Fig. 7), combines the drug and sometimes an osmotic agent in a monolithic core and delivers the drug in solution (102). The mass delivery rate with time (dm/dt) of the drug solution is described by equation 4, where L_p is the hydraulic permeability of the membrane, α is the membrane reflection coefficient, $\Delta\pi$ is the osmotic pressure gradient, ΔP is the hydrostatic back pressure, A is the area of the membrane, C is the dissolved concentration of the drug, and h is the membrane thickness.

$$\frac{dm}{dt} = \frac{L_p(\alpha\Delta\pi - \Delta P) AC}{h}$$

(4)

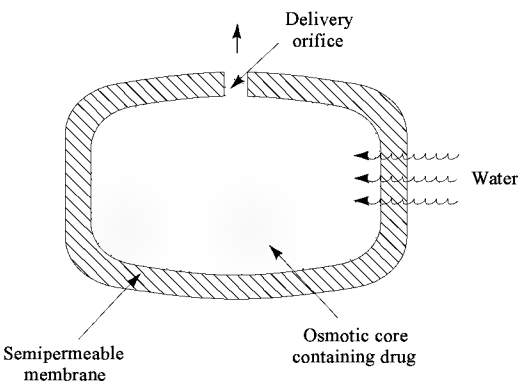


Fig. 7. A basic elementary osmotic pump.

Courtesy of ALZA Corp.

The delivery port can be sized sufficiently large that the back pressure is negligibly small. Equation 4 then reduces to equation 5, where k represents the osmotic water permeability of the membrane and S represents the solubility of the drug at saturated concentration.

$$\frac{dm}{dt} = \frac{k\Delta\pi AS}{h}$$

(5)

This equation describes the steady-state, or zero-order, release of the drug. When the drug completely dissolves, its concentration within the system begins to dilute, and the release rate follows a parabolic decline with time (102). Acutrim (ALZA Corp.), delivering phenylpropanolamine hydrochloride [154-41-6] for appetite suppression, is an example of an elementary osmotic pump.

Another type of osmotic device has the osmotic agent and the drug in separate layers (Fig. 8). Incoming water causes the osmotic layer to hydrate and expand and the drug layer gradually to go into solution or suspension. The dissolved or suspended drug is released from the system orifice(s). The net effect is to release the drug at a zero-order rate. This Push–Pull system can release a variety of compounds, ranging from highly soluble drugs to insoluble drugs suspended in osmotic hydrogel carriers or in carriers that solvate at body temperature. Procardia XL, delivering nifedipine for angina and hypertension, is a Push–Pull osmotic system.

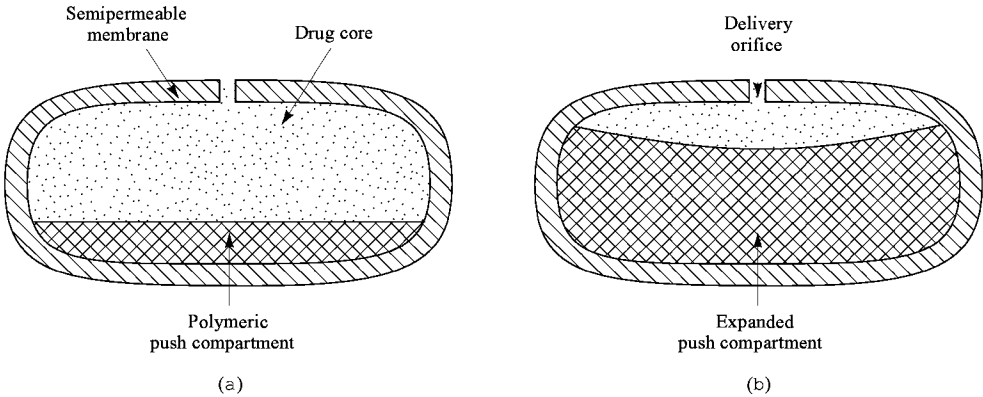


Fig. 8. Example of a Push–Pull osmotic pump where (a) represents the pump before operation, and (b), during operation.

Courtesy of ALZA Corp.

Two osmotic systems have been developed primarily for research. The Osmet delivery module is used for clinical studies, and the Alzet mini-osmotic pump (Fig. 9) is used for nonclinical studies. Both systems are fabricated without drugs and can be filled by the researcher with various drugs in solution or suspension. The drug reservoir is a deformable, elastomeric chamber coated with an osmotic layer that is overcoated with a semipermeable membrane. As the systems imbibe water, a hydrostatic pressure is generated between the membrane and the wall of the drug chamber, deforming the chamber and slowly releasing the drug. The Osmet module has been used to explore the effects on drug actions of rate-controlled oral, vaginal, or rectal drug delivery and is designed to deliver a drug over an 8-, 12-, or 24-h period. The Alzet system is a miniaturized, implantable osmotic pump that can deliver a drug to a body cavity, blood vessel, or local tissue site. The reservoir volume of commercially available Alzet systems ranges from 100 to 2000 μL , and delivery durations range from days to weeks (103).

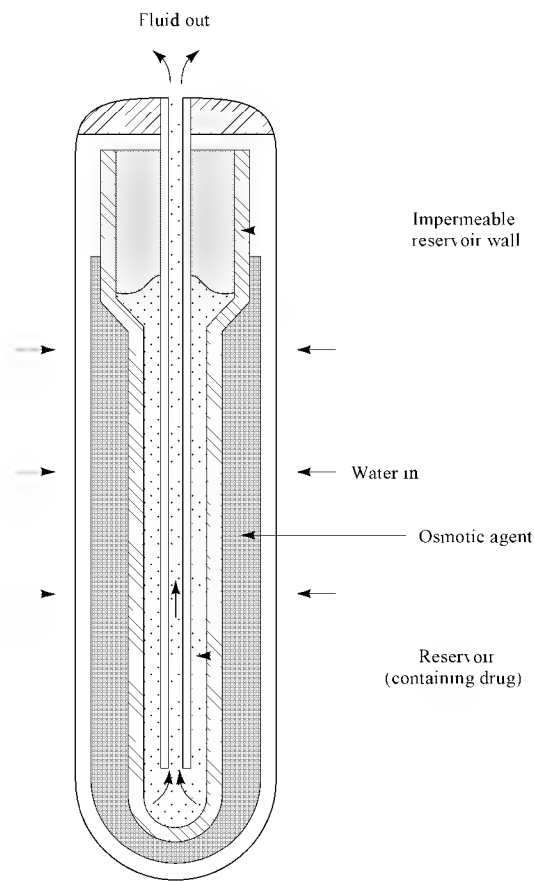


Fig. 9. Mini-osmotic pump.

Courtesy of ALZA Corp.

The Push-Melt osmotic system (Fig. 10) (104), developed primarily for zero-order delivery of drugs to the rumina of cattle, consists of a surrounding semipermeable membrane, typically cellulose ester polymers plus plasticizers (105); an osmotic tablet, such as sodium chloride [7647-14-5] dispensed in sodium polyacrylate [9003-04-7]; a separating layer; a drug dispersed in a thermoresponsive carrier; and an iron weight that also contains the delivery orifice. The iron weight increases the density of the system so that it is retained in the rumen for more than 120 days. The thermoresponsive carrier melts at body temperature, allowing the drug to be released from the system as the osmotic tablet imbibes water. This dosage form has been developed to deliver the anthelmintic drug ivermectin [70288-86-7] (IVOMEC SR bolus) and the nutritional supplement sodium selenite [10102-18-8] (Dura Se bolus), both from ALZA Corp.

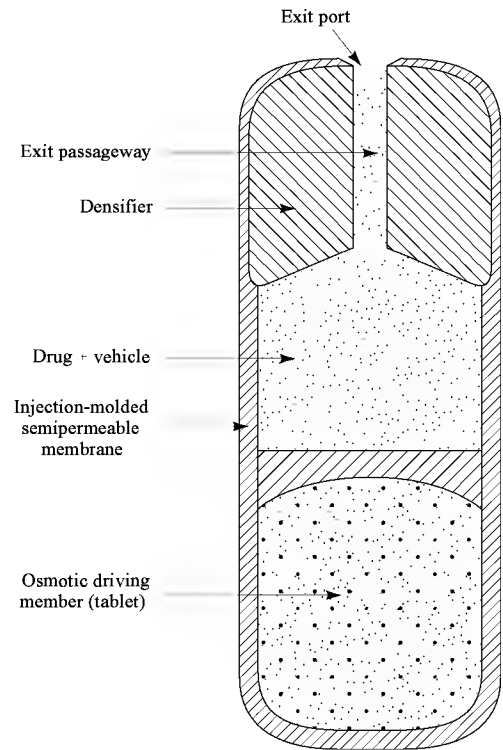


Fig. 10. Structure of Push-Melt osmotic pump, developed primarily for veterinary therapeutic applications.

Courtesy of ALZA Corp.

A drug-dedicated osmotic implant for human and veterinary use has been developed to deliver hormones, peptides, and proteins that are digested or rendered inactive after oral administration (106).

pH Sensitivity. Dosage forms that employ pH sensitivity to trigger the release of active agents have been designed with both physical and chemical mechanisms. Systems with physical mechanisms are discussed here; a description of chemical mechanisms, such as pH-sensitive hydrogels, is available (107).

Delivery systems that respond to changes in pH have been known to the pharmaceutical industry for more than a century. The pH-sensitive enteric coating is probably the oldest controlled-release technology. Unna introduced an enteric tablet coating based on keratin in 1884 (108). Enteric coatings are used primarily to protect the gastric mucosa from local irritation or to ensure that tablets do not dissolve until they reach the intestine.

Enteric coatings are generally formulated from anionic polymers with pendent carboxyl groups and typically have a pK_a of 4 to 6. In the low pH gastric fluids (about two units below the pK_a), only 1% of the carboxyl groups ionize; 99% of the carboxyl groups are protonated, and the carboxyl groups can form hydrogen bonds with each other or with other portions of the polymer. The insoluble polymer film thus retains its integrity and provides a barrier to moisture. When the dosage form reaches the higher pH intestinal fluids, ionization of the pendent groups increases and the enteric material dissolves (109). The extent of this ionization as a function of media pH can be described by the Henderson-Hasselbach equation (110):

$$pH = pK_a + \log \frac{\text{ionized}}{\text{nonionized}} \tag{6}$$

The most commonly used polymers are cellulose acetate phthalate [9004-38-0] (CAP), poly(vinyl acetate phthalate) [34481-48-6] (PVAP), hydroxypropylmethyl-cellulose phthalate [71138-97-1] (HPMCP), and polymethacrylates (111) (see CELLULOSE ESTERS). Acrylate copolymers are also available (112). Figure 11 shows the dissolution behavior of some commercially available enteric materials. Some manufacturers supply grades designed to dissolve at specific pH values with increments as small as 0.5 pH unit (113).

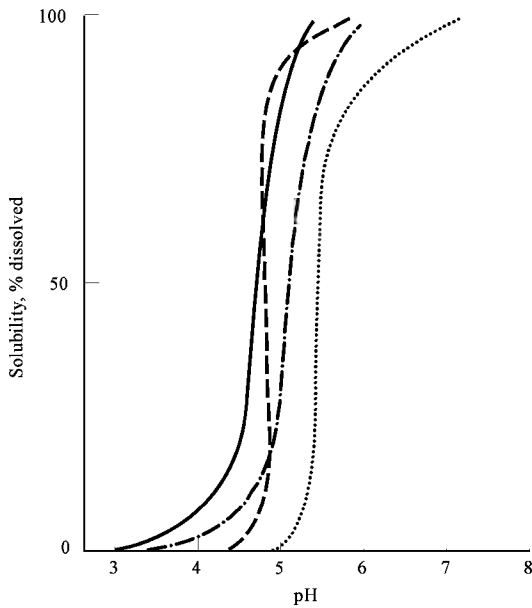


Fig. 11. Dissolution as a function of pH of various enteric polymers. (—) represents HPMCP (HP-50); (---), HPMCP (HP-55); (···), PVAP; (- · - ·), CAP.

Courtesy of Colorcon, Inc., West Point, Pa.

Other dosage forms can be designed to use pH sensitivity. Examples include microcapsules or liposomes that release drugs by changes in pH (114) and may be used to treat tumor cells, which reportedly have a pH substantially lower than that of healthy tissues; and a dosage form incorporating pH-sensitive electrodes in a feedback system with mechanical pumps (115) that continuously monitors gastric pH values with a nasogastric electrode and intravenously infuses appropriate levels of antiulcer medication in response to changes in intragastric pH.

Vapor Pressure. The Shiley Infusaid implantable infusion pump utilizes energy stored in a two-phase fluorinated hydrocarbon fluid. The pump consists of a refillable chamber that holds the drug and a chamber that holds the fluid. The equilibrium vapor pressure of the fluid, a constant 60 kPa (450 mm Hg), compresses the bellows, pumping the drug through a bacterial filter, a capillary flow restrictor, and an infusion cannula to the target body site (56,116).

Emerging Applications of Controlled-Release Technology

Several application areas of controlled-release technology will be increasingly important. Systems being investigated or developed employ many of the mechanisms discussed earlier.

Patterned Delivery. Although steady serum–drug concentrations are the preferred mode of therapy for many drugs, patterned delivery of some compounds, including the peptides calcitonin [9007-12-9], growth hormone [9002-72-6], and corticotropin-releasing hormone [9015-71-8], mimics the body's own secretion patterns or responses to certain compounds or diseases. Osmotic pumps can be programmed to deliver drugs in a variety of patterns. In one patterned delivery system for nocturnal asthma, an elementary osmotic pump releases a delayed pulse of salbutamol [18559-94-9], superimposed over a steady delivery rate, approximately 7 h after dosing (Fig. 12). For cattle, the Synanthic Multidose system delivers five pulses of the anthelmintic oxfendazole [53716-50-0] over a 4-month period. The system consists of a central magnesium rod attached to a steel and weight. Plastic segments containing the drug are sequentially positioned on the rod. In the rumen, the rod erodes as a result of electrolytic action, and the plastic segments are delivered at 3-wk intervals (118,119). Patterned delivery is also possible with electrically assisted transdermal systems that apply an intermittent, and sometimes varying, electrical charge to enhance drug permeation at programmed intervals.

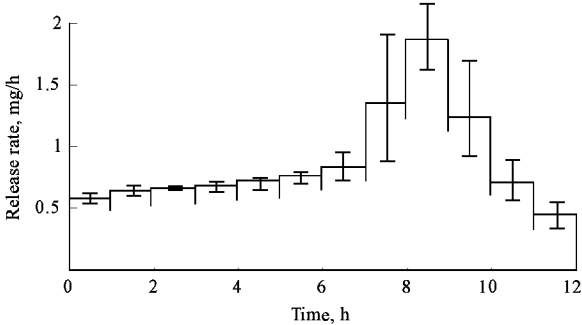


Fig. 12. Profile of patterned delivery of salbutamol, for nocturnal asthma, from an elementary osmotic pump. A delayed pulse of salbutamol, superimposed over a steady delivery rate, is released approximately 7 h after dosing (117).

Courtesy of ALZA Corp.

Targeted Delivery. By delivering drugs to the specific cells or organs where action is required, targeted dosage forms may enhance therapeutic outcome and avoid many toxic effects. Colon-targeted systems, for example, enable a drug to reach its destination without being digested in the stomach. One approach to targeted delivery is to couple a pharmacological agent to a macromolecular vector specifically taken up by target cells. Vectors include antibodies, nanoparticles, polysaccharides, synthetic polypeptides, liposomes, synthetic and natural cells, polymer-based microcapsules, albumin, and glycoprotein conjugates (120–122). Recently, companies have begun developing systems to circumvent the blood–brain barrier and target the brain. CytRx has patented a copolymer for treatment of tumors (123); Pharmatec US Neurex claims to be able to deliver neuropeptides to the brain with a patented carrier system (124); and Gynex claims to deliver estradiol to the brain (125).

Triggered and Closed-Loop Delivery. Triggered delivery systems release a drug in response to a signal from the body, for example, the presence of a substance to be controlled, such as morphine, or from the patient. Triggered delivery of the morphine antagonist naltrexone [16590-41-3] has been explored by several groups. A device containing naltrexone dispersed in a biodegradable polymer core, eg, a partially esterified copolymer of methyl vinyl ether [107-25-5] and maleic anhydride [108-31-6] that is stable at low pH but erodes rapidly at pH 7.4 (a physiologic pH), has been described (126) (Fig. 13). A protective coating (an acidic hydrogel or a hydrophobic compound) surrounds the polymer, and a microporous membrane surrounds the entire device. An antibody-blocked enzyme between the protective coating and the membrane is freed by the entry of morphine from the body. The free enzyme degrades the protective coating around the core and exposes it to the body's higher pH, allowing the release of naltrexone (126–129).

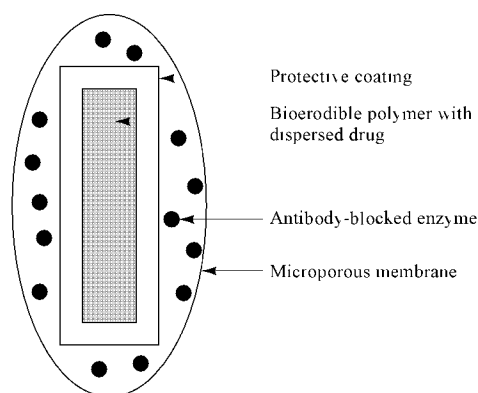


Fig. 13. Proposed triggered delivery system with the drug dispersed in a rate-controlled polymer. A protective coating surrounds the polymer, and a microporous membrane surrounds the entire device.

Courtesy of Butterworth-Heinemann Ltd., Oxford, UK.

The ultimate controlled-release device, a closed-loop system, delivers a metered amount of a drug each time it encounters a signal from the body, for example, the plasma concentration of a metabolite or a substance such as glucose [50-99-7], until the system is depleted (8,128–130). Closed-loop insulin delivery has been the focus of much research. Rapidly changing blood glucose levels require system response times of 15 min or less (131,132). One approach envisions a hydrogel polymer membrane containing pendent tertiary amine groups and entrapped glucose oxidase [9001-37-0]. Glucose oxidase catalyzes the reaction of glucose and oxygen to produce gluconic acid and hydrogen peroxide [7722-84-1]. As the membrane pH decreases, the pendent amine groups are ionized, resulting in membrane swelling and increased membrane permeability to insulin. The principal hydrogel types investigated are based on 2-hydroxyethyl methacrylate [868-77-9], *N,N*-dimethylaminoethyl methacrylate [2867-47-2], and polyacrylamide [9003-05-8] (119).

Another approach utilizes the plant lectin concanavalin A (Con A) [11028-71-0], which has a high binding affinity for specific saccharides. Con A is immobilized on Sepharose [9012-36-6] beads or cross-linked to form a gel. Glycosylated insulin (*p*-succinyl aminophenyl glucopyranoside insulin) is bound to the sites on the Con A. Con A–glycosylated insulin is then placed in a microporous synthetic membrane pouch. When glucose diffuses through the membrane, it displaces the bound glycosylated insulin, which diffuses out through the membrane (133,134).

A mechanochemical pump being developed incorporates a glucose-sensitive hydrogel. As glucose diffuses into the hydrogel in one chamber, it reacts with glucose oxidase, changing the pH of the hydrogel and causing it to swell. The swelling exerts pressure against a diaphragm, pressurizing two adjoining chambers and opening a valve to release insulin. The pH-sensitive hydrogels investigated to date have been linear copolymers of *N,N*-diethylaminoethyl methacrylate hydrochloride [2421-44-5] and nonionic methacrylates (135,136) (see METHACRYLIC POLYMERS).

BIBLIOGRAPHY

"Pharmaceuticals, Controlled Release," in *ECT* 3rd ed., Vol. 17, pp. 290–310, by H. Benson, B. Harley, and E. E. Schmitt, ALZA Corp.

1. M. S. Gately, J. R. *Coll. Gen. Pract.* **16**, 39–44 (1968).
2. J. Skelly, *Oral Controlled Release Products: Therapeutic and Biopharmaceutic Assessment*, Wissenschaftliche Verlagsgesellschaft mbH, Stuttgart, Germany, 1990, pp. 175–194.
3. *Physicians' Desk Reference*, 45th ed., Medical Economics Co., Inc., Oradell, N.J., 1991.
4. J. Urquhart, *Controlled-Release Pharmaceuticals*, American Pharmaceutical Assn., Washington, D.C., 1981, pp. 1–48.
5. J. C. Johnson, ed., *Sustained Release Medication*, Noyes Data Corporation, Park Ridge, N.J., 1980.
6. J. R. Robinson, ed., *Sustained and Controlled Release Drug Delivery Systems*, Marcel Dekker, Inc., New York, 1978, p. iii.
7. R. D. Bagnall, *Biomater. Med. Devices Artif. Organs* **5**, 355–359 (1977).
8. F. Theeuwes, *Novel Drug Delivery and Its Therapeutic Application*, John Wiley & Sons, Ltd., West Sussex, England, 1989, pp. 323–340.
9. H. Staudinger, in Ref. 2, pp. 83–98.
10. B. M. Silber, M. Bialer, and A. Yacobi, *Controlled Drug Delivery: Fundamentals and Applications*, 2nd ed., Marcel Dekker, Inc., New York, 1987, pp. 213–251.
11. F. E. Yates, H. Benson, R. Buckles, J. Urquhart, and A. Zaffaroni, *Advances in Biomedical Engineering*, Academic Press, New York, 1975, pp. 2–34.
12. R. Langer, *Science* **249**, 1527–1533 (1990).
13. V. H. L. Lee, ed., *Peptide and Protein Drug Delivery*, Marcel Dekker, Inc., New York, 1991.
14. S. S. Davis, L. Illum, and E. Tomlinson, eds., *Delivery Systems for Peptide Drugs*, Plenum Press, New York, 1986.
15. I. E. Walter-Sack, in Ref. 2, pp. 117–137.
16. R. Baker, *Controlled Release of Biologically Active Agents*, John Wiley & Sons, Inc., New York, 1987, pp. 221–265.
17. S. S. Davis, *J. Control. Rel.* **2**, 27–38 (1985).
18. C. A. Sninsky and co-workers, *Ann. Intern. Med.* **115**, 350–354 (1991).
19. L. A. Christensen and co-workers, *Aliment. Pharmacol. Therap.* **4**, 523–533 (1990).
20. P. J. Watts, M. C. Davies, and C. D. Melia, *J. Control. Rel.* **16**, 311–318 (1991).

21. F. Theeuwes, P. Wong, T. Burkoth, and D. Fox, *Colonic Drug Absorption and Metabolism*, Marcel Dekker, Inc., New York, in press.
22. F. Theeuwes, S. I. Yum, R. Haak, and P. Wong, *Temporal Control of Drug Delivery*, Vol. 618, *Annals of The New York Academy of Sciences*, N.Y.A.S., New York, 1991, pp. 428–440.
23. Y. W. Cho and M. Flynn, *Lancet* **2**, 1518–1519 (1989).
24. B. Lippold, in Ref. 2, pp. 39–57.
25. M. A. Longer and J. R. Robinson, *Remington's Pharmaceutical Sciences*, Vol. 18, Philadelphia College of Pharmacy and Science, Mack Printing Co., Easton, Pa., 1990, pp. 1676–1693.
26. H.-W. Hui, J. R. Robinson, and V. H. L. Lee, *Controlled Drug Delivery: Fundamentals and Applications*, 2nd ed., Marcel Dekker, Inc., New York, 1987, pp. 373–432.
27. L. Krywczynski and D. P. Brozyna, *Extended-Release Dosage Forms*, CRC Press, Inc., Boca Raton, Fla., 1987, pp. 1–19.
28. K. Morimoto and co-workers, *J. Pharm. Pharmacol.* **42**, 720–722 (1990).
29. D. D. Breimer, A. G. de Boer, and L. G. J. de Leede, *J. Control. Rel.* **2**, 39–46 (1985).
30. C. M. Maloney, R. K. Kesner, G. Klein, and J. Bockenstette, *Am. J. Hospice Care*, 34–35 (1989).
31. L. Cole, C. D. Hanning, S. Robertson, and K. Quinn, *Br. J. Clin. Pharmacol.* **30**, 781–786 (1990).
32. C. W. Twiston-Davies, M. I. Goodwin, and P. J. Baxter, *J. Bone J. Surg. [Br]* **72-B**, 510–511 (1990).
33. F. Theeuwes, P. Wong, S. I. Yum, *Encyclopedia of Pharmaceutical Technology*, Vol. 4, Marcel Dekker, Inc., New York, 1991, pp. 303–348.
34. Y. W. Chien and C.-S. Lee, *Controlled-Release Technology: Pharmaceutical Applications*, ACS Symposium Series 348, American Chemical Society, Washington, D.C., 1987, pp. 281–300.
35. M. Pedersen and M. R. Rassing, *Drug Dev. Ind. Pharm.* **16**, 2015–2030 (1990).
36. M. Addy and D. R. Fugit, *Clin. Materials* **4**, 271–284 (1989).
37. M. Friedman and D. Steinberg, *Pharm. Res.* **7**, 313–317 (1990).
38. P. B. Deasy, A. E. M. Collins, D. J. MacCarthy, and R. J. Russell, *J. Pharm. Pharmacol.* **41**, 694–699 (1989).
39. Ref. 3, pp. 1079–1080.
40. W. A. Lee and J. P. Longenecker, *BioPharm*, 30–37 (Apr. 1988).
41. R. D. Ennis, L. Borden, and W. A. Lee, *Pharm. Res.* **7**, 468–475 (1990).
42. S. S. Davis, L. Illum, D. Burgess, J. Ratcliffe, and S. N. Mills, in Ref. 34, pp. 201–213.
43. F. W. H. M. Merkus, *Rate-Controlled Drug Administration and Action*, CRC Press, Inc., Boca Raton, Fla., 1986, pp. 15–47.
44. V. H. L. Lee, *J. Ocul. Pharmacol.* **6**, 157–164 (1990).
45. G. Hecht, R. E. Roehrs, E. R. Cooper, J. W. Hiddemen, and B. F. Van Duzee, *Modern Pharmaceutics*, 2nd ed., Marcel Dekker, Inc., New York, 1990, pp. 539–603.
46. D. L. Middleton, S.-H. S. Leung, and J. R. Robinson, *Bioadhesive Drug Delivery Systems*, CRC Press, Inc., Boca Raton, Fla., 1990, pp. 179–202.
47. V. H. K. Li, J. R. Robinson, and V. H. L. Lee, *Controlled Drug Delivery: Fundamentals and Applications*, Marcel Dekker, Inc., New York, 1987, pp. 3–94.
48. V. H. L. Lee and J. R. Robinson, *J. Ocul. Pharm.* **2**, 67–108 (1986).
49. J. Urquhart, *Ophthalmic Drug Delivery Systems*, Academy of Pharmaceutical Science, Washington, D.C., 1980, pp. 105–118.
50. *Physicians' Desk Reference for Ophthalmology*, 19th ed., Medical Economics, Co., Oradell, N.J., 1991, p. 255.
51. S. D. Schwartz and co-workers, *Am. J. Ophthalmol.* **109**, 701–704 (1990).
52. R. Bawa and M. Nandu, *Biomaterials* **11**, 724–728 (1990).
53. M. Nomura and co-workers, *J. Pharm. Pharmacol.* **42**, 292–294 (1990).
54. N. Washington, C. G. Wilson, and C. Washington, *Physiological Pharmaceutics: Biological Barriers to Drug Absorption*, Ellis Horwood Ltd., Chichester, England, 1989, pp. 155–178.
55. I. Gonda, *Crit. Rev. Ther. Drug Carrier Syst.* **6**, 273–313 (1990).
56. T. A. McCalden, *Advanced Drug Delivery Reviews* **5**, 253–263 (1990).
57. H. Katayama, Z. Sun, R. W. Niven, and F. Rypacek, *Am. Chem. Soc. Abstracts* **72** (1990).
58. A. Adjei and J. Garren, *Pharm. Res.* **7**, 565–569 (1990).
59. P. D. Darney, C. M. Klaisle, S. Tanner, and A. M. Alvarado, *Curr. Probl. Obstet. Gynecol. Fertil.*, 87–125 (May June 1990).
60. T. Luukkainen, P. Lahteenmaki, and J. Toivonen, *Ann. Med.* **22**, 85–90 (1990).
61. I. Z. MacKenzie and A. V. G. Taylor, *Lancet* **562** (Sept. 1, 1990).
62. C. Read, *New Sci.* **28** (Mar. 11, 1989).
63. F. Elzik and co-workers, *J. Clin. Endocrinol. Metab.* **63**, 29–35 (1986).
64. P. Tyle, ed., *Drug Delivery Devices: Fundamentals and Applications*, Marcel Dekker, Inc., New York, 1988.
65. T. T. Rohde, H. Buchwald, and P. J. Blackshear, *Drug Delivery Devices: Fundamentals and Applications*, Marcel Dekker, Inc., New York, 1988, pp. 235–260.
66. J. Urquhart, speech, *New Technologies in the Health Care Industry Conference*, Brussels, Belgium, Oct. 13–14, 1980.
67. K. W. Leong, *Polymers for Controlled Drug Delivery*, CRC Press, Inc., Boca Raton, Fla., 1991, pp. 127–148.
68. M. R. Brophy and P. B. Deasy, *Encyclopedia of Pharmaceutical Technology*, vol. 2, Marcel Dekker, Inc., New York, 1990, pp. 1–25.
69. R. Jalil, *Drug Dev. Ind. Pharm.* **16**, 2353–2367 (1990).
70. R. J. Linhardt, *Controlled Release of Drugs: Polymers and Aggregate Systems*, VCH Publishers, New York, 1989, pp. 53–95.
71. J. Heller, *Crit. Rev. Ther. Drug Carrier Syst.* **1**, 39–90 (1984).
72. S. Yolles, J. E. Eldridge, and J. H. R. Woodland, *Polym. News* **1**, 9–15 (1970).
73. Ref. 3, p. 2211.
74. I. J. Tarr, ed., *Pharmaprojects*, Vol. II, PJB Publications, Ltd., United Kingdom, May 1991, p. a704.
75. Ref. 3, pp. 1078–1079.
76. L. M. Sanders, *Eur. J. Drug Metab. Pharmacokinet.* **15**, 95–102 (1990).
77. L. Sanders, R. Burns, K. Vitale, and P. Hoffman, *Proc. Int. Symp. Control. Rel. Bioact. Mater.* **15**, 62–63 (1988).
78. R. L. Dunn and co-workers, *Polym. Prepr.* **31**, 189–191 (1990).
79. *Advancing Ophthalmic Drug Therapy*, InSite Vision company promotional publication, Alameda, Calif., 1989.
80. T. Higuchi, *J. Pharm. Sci.* **50**, 874–875 (1961).
81. D. R. Paul and S. K. McSpadden, *J. Membr. Sci.* **1**, 33–48 (1976).
82. V. A. Place and K. C. Nichols, in Ref. 22, pp. 441–449.
83. T. H. Ferguson, G. F. Needham, and J. F. Wagner, *J. Control. Rel.* **8**, 45–54 (1988).
84. R. Langer and J. Folkman, *Nature* **263**, 797–800 (1976).
85. H. M. Leeper and co-workers, *Rubber Chem. Technol.* **50**, 969–980 (1977).
86. S. Yum, R. Buckles, and H. Leeper, *J. Elastomers Plast.* **10**, 340–354 (1978).
87. **U.S. Pat. 3,993,069** (Nov. 23, 1976), R. Buckles, H. Leeper, S. I. Yum, and A. S. Michaels (to ALZA Corp.).
88. D. Winchell and J. Tune, *Drug Delivery Devices: Fundamentals and Applications*, Marcel Dekker, Inc., New York, 1988, pp. 213–234.

89. R. R. Burnette and D. Marrero, *J. Pharm. Sci.* **75**, 738–743 (1986).

90. R. R. Burnette and B. Ongpipattanakul, *J. Pharm. Sci.* **77**, 132–137 (1988).

91. A. J. Bard and L. R. Faulkner, *Electrochemical Methods: Fundamentals and Applications*, John Wiley & Sons, Inc., New York, 1980, p. 120.

92. G. B. Kasting and J. C. Keister, *J. Control. Rel.* **8**, 195–210 (1989).

93. K. Tojo, *J. Chem. Eng. Jpn.* **22**, 512–518 (1989).

94. R. R. Burnette, *Transdermal Drug Delivery: Developmental Issues and Research Initiatives*, Marcel Dekker, Inc., New York, 1989, pp. 247–291.

95. M. J. Pikal, *Phar. Res.* **7**, 118–126 (1990).

96. P. Tyle and B. Kari, *Drug Delivery Devices: Fundamentals and Applications*, Marcel Dekker, Inc., New York, 1988, pp. 421–454.

97. B. R. Meyer and co-workers, *Clin. Pharmacol. Ther.* **44**, 607–612 (1988).

98. B. R. Meyer and co-workers, *Am. J. Med. Sci.* **297**, 321–325 (1989).

99. Y. W. Chien, P. Lelawongs, O. Siddiqui, Y. Sun, and W. M. Shi, *J. Control. Rel.* **13**, 263–278 (1990).

100. S. Kumar and co-workers, *Proc. Int. Symp. Control. Rel. Bioact. Mater.* **17**, 435–436 (1990).

101. L. R. Tamburrini, M. DiMonte, and P. Sfredde, *Technologiebiomediche* **7**, 60–64 (1987).

102. F. Theeuwes, *J. Pharm. Sci.* **64**, 1987–1991 (1975).

103. Database of Alzet Publications, ALZA Corp., Palo Alto, Calif., 1992.

104. B. Eckenhoff, F. Theeuwes, and J. Urquhart, *Pharm. Tech.*, 34–44 (Jan. 1981).

105. **U.S. Pat. 4,731,122** (Mar. 15, 1988), R. Cortese, J. C. Wright, J. B. Eckenhoff, and D. L. Rivera (to ALZA Corp.).

106. **U.S. Pat. 5,037,420** (Aug. 6, 1991), J. A. Magruder, J. B. Eckenhoff, R. Cortese, J. C. Wright, and J. R. Peery (to ALZA Corp.).

107. J. Heller, A. C. Chang, G. Rodd, G. M. Grodsky, *J. Control. Rel.* **13**, 295–302 (1990).

108. M. A. Longer and J. R. Robinson, *Remington's Pharmaceutical Sciences*, Vol. 18, Philadelphia College of Pharmacy and Science, Mack Printing Co., Easton, Pa., 1990, p. 1681.

109. W. G. Chambliss, *Pharm. Technol.* **7**, 124–140 (1983).

110. A. Martin, J. Swarbrick, and A. Cammarata, *Physical Pharmacy*, 3rd ed., Lea & Febiger, Philadelphia, 1983, pp. 222–223.

111. J. Sjogren, *Rate Control in Drug Therapy*, Churchill Livingstone, Edinburgh, 1985, pp. 38–47.

112. Rohm Pharma Eudragit L100 *Technical Bulletin Info L-4/e* (May 1989).

113. Shin-Etsu *HPMCP Technical Bulletin 85.6.1000 NP* (1985).

114. M. B. Yatvin, W. Kreutz, B. A. Horwitz, and M. Shinitzky, *Science* **210**, 1253–1255 (1980).

115. H. Merki and co-workers, *Eur. J. Gastroenterol. Hepatol.* **3**, 9–13 (1991).

116. J. Johnston, S. Reich, A. Bailey, and J. Sleutz, *Ann. N.Y. Acad. Sci.*, **531**, 57–65 (1988).

117. **U.S. Pat. 4,751,071** (June 14, 1988), P. R. Magruder and co-workers (to ALZA Corp.).

118. G. L. Zimmerman and E. P. Hoberg, *Parasitol. Today* **4**, 55–56 (1988).

119. J. A. Bogan, J. A. Armour, K. Bairden, and E. A. Galbraith, *Vet. Rec.* **121**, 280 (1987).

120. E. Tomlinson, *Site-Specific Drug Delivery*, John Wiley & Sons Ltd., Chichester, England, 1986, pp. 1–26.

121. L. Fiume, C. Busi, A. Mattioli, and G. Spinosa, *CRC Critical Reviews in Therapeutic Drug Carrier Systems* **4**, 265–284 (1988).

122. C. A. Hunt and R. D. MacGregor, *Topics in Pharmaceutical Sciences 1987*, Elsevier Science Publishers B.V., Amsterdam, 1987, 409–420.

123. *Clinica (World Medical Device and Diagnostic News)*, PJB Publications Ltd., London, 1990, **384**, 5.

124. *Scrip (World Pharmaceutical News)*, vol. 1407, PJB Publications Ltd., London, 1989, p. 15.

125. N. Bodor, *Biological Approaches to the Controlled Delivery of Drugs*, Vol. 507, *Annals of the New York Academy of Sciences*, N.Y.A.S., New York, 1987, pp. 289–306.

126. J. Heller, S. H. Pangburn, and K. V. Roskos, *Biomaterials* **11**, 345–350 (1990).

127. K. V. Roskos, C. L. English, D. R. Friend, and J. Heller, *Proc. Int. Symp. Control. Rel. Bioact. Mater.* **16**, 69 (1989).

128. J. Heller, *J. Controlled Rel.* **8**, 111–125 (1988).

129. C. G. Pitt, *Pharmacy Int.* **7**, 88–91 (1986).

130. J. Kost, ed., *Pulsed and Self-Regulated Drug Delivery*, CRC Press, Inc., Boca Raton, Fla., 1990.

131. J. Heller, in Ref. 130, pp. 93–108.

132. G. Albin, T. A. Horbett, and B. D. Ratner, in Ref. 130, pp. 159–185.

133. L. A. Seminoff and S. W. Kim, in Ref. 130, pp. 187–199.

134. S. W. Kim and co-workers, *J. Controlled Rel.* **11**, 193–201 (1990).

135. R. A. Siegel and B. S. Firestone, *J. Controlled Rel.* **11**, 181–192 (1990).

136. J. M. Cornejo-Bravo and R. A. Siegel, *Proc. Int. Symp. Control. Rel. Bioact. Mater.* **17**, 71 (1990).

General References

R. Baker, *Controlled Release of Biologically Active Agents*, John Wiley & Sons, Inc., New York, 1987, pp. 221–265.

Y. W. Chien, *Encyclopedia of Pharmaceutical Technology*, vol. 3, Marcel Dekker, Inc., New York, 1990, pp. 281–313.

R. Duncan and L. W. Seymour, *Controlled Release Technologies: A Survey of Research and Commercial Applications*, Elsevier Advanced Technology, Oxford, UK, 1989.

U. Gundert-Remy and H. Müller, eds., *Oral Controlled Release Products: Therapeutic and Biopharmaceutic Assessment*, Wissenschaftliche Verlagsgesellschaft mbH, Stuttgart, Germany, 1990.

L. Illum and S. S. Davis, eds., *Polymers in Controlled Drug Delivery*, IOP Publishing Ltd., Bristol, England, 1987.

R. L. Juliano, ed., *Biological Approaches to the Controlled Delivery of Drugs*, vol. 507, *Annals of the New York Academy of Sciences*, N.Y.A.S., New York, 1987.

L. Krywczynski and D. P. Brozyna, *Extended-release Dosage Forms*, CRC Press, Inc., Boca Raton, Fla., 1987, pp. 1–19.

P. I. Lee and W. R. Good, eds., *Controlled-Release Technology: Pharmaceutical Applications, ACS Symposium Series 348*, American Chemical Society, Washington, D.C., 1987.

J. W. McGinity, ed., *Aqueous Polymeric Coatings for Pharmaceutical Dosage Forms*, Marcel Dekker, Inc., New York, 1989.

J. R. Robinson and V. H. L. Lee, eds., *Controlled Drug Delivery: Fundamentals and Applications*, 2nd ed., Marcel Dekker, Inc., New York, 1987.

M. Rosoff, ed., *Controlled Release of Drugs: Polymers and Aggregate Systems*, VCH Publishers, Inc., New York, 1989.

M. H. Rubinstein, *Pharmaceutical Technology: Controlled Drug Release*, vol. 1, Ellis Horwood, Ltd., Chichester, England, 1987.

H. A. J. Struyker-Boudier, ed., *Rate-Controlled Drug Administration and Action*, CRC Press, Inc., Boca Raton, Fla., 1986.

P. Tyle, ed., *Drug Delivery Devices: Fundamentals and Applications*, Marcel Dekker, Inc., New York, 1988.

A. Yacobi and E. Halperin-Walega, eds., *Oral Sustained Release Formulations: Design and Evaluation*, Pergamon Press, New York, 1988.

Journals

Advanced Drug Delivery Reviews, Elsevier, Amsterdam.

Journal of Controlled Release, Controlled Release Society, Elsevier, Amsterdam.

David Edgren
Harold Leeper
Kirstin Nichols
Jeremy Wright
ALZA Corporation

CONVEYING

Conveying is a term used for the transport of bulk solids. Bulk solids conveyors are in large part made up of components that are dimensionally standardized, but fabricated in varying classes of construction to meet operating conditions ranging from light-duty intermittent operation to heavy-duty continuous operation. Conveyors are volumetric machines that transport material fed to them at a controlled rate. This is in contrast to solids feeders that operate under a solids head and control the volumetric flow. By appropriate changes in geometry, most conveyors can be used as volumetric feeders. By addition of appropriate weight sensing and control modules, conveyors can also be used as gravimetric feeders. The selection of the type of conveyor for a specific application is dependent first on the required capacity; second, on the conveying path, ie, horizontal, vertical, or a combination of both; and third, on the handling characteristics of the material.

Characterization of Bulk Material

The basis of all bulk conveyor engineering is the precise definition and accurate classification of materials according to individual characteristics under a specific combination of handling conditions (1). Since the late 1960s there has been an extraordinary growth in research into the fundamental properties and behavior of particulate solids. However, as of this writing, it is not possible to predict the handling behavior of a bulk solids material relevant to conditions in a specific conveyor, merely on the basis of the discrete particle properties.

The laws governing gravity flow within channels in silos and bins can be determined by treating the bulk solids as a continuum, where the properties are a continuous function (2). The mass can then be divided indefinitely without losing these defining properties. As a result, the Jenike flow theories, shear tests, and design procedures have been developed that make it possible to design bulk solids storage bins for reliable and predictable flow, with a high degree of success. Boundary conditions within bulk handling systems are considerably different from those within silos, however, so this analysis has not yet produced a useful general correlation for predicting behavior in conveying situations. Although intended primarily for bin design, the Jenike shear tester is often used to measure *relative* flowability of a powder or granular material. The test procedure also includes measurement of the coefficient of friction between a moving bed of material and a stationary wall surface that replicates the wall surface in a silo. The friction values obtained in this manner can be useful for some, but not all, conveyor design situations, because the relative rubbing velocities (about 3 mm min) used in the test are much lower than those normally encountered in most dry bulk and powder conveying systems (see also POWDERS, HANDLING).

Industry practice is to describe flow behavior by descriptive terms derived by combining data from empirical bench-top tests and observations of actual conveying systems in operation. This information is then used first, as a guide in the selection of the particular type of conveyor, and second, for determining the design features required for the particular application. Collections of most of this information can be found in the guides published by manufacturers' engineering associations in various countries. These guides contain extensive listings of measured and observed handling characteristics for specific materials. Typical are the publications of the Conveyor Equipment Manufacturers Association (CEMA) (1) in the United States, used primarily for mechanical conveying.

The CEMA publication gives terminology, definitions, and test procedures for measuring and describing 37 characteristics of a material. Each measured characteristic or property is assigned an alpha-numeric classification code designation. Thus the *CEMA Classification Code* provides a convenient method of communication between users and manufacturers of conveying equipment. Other countries have developed similar standardized tests, most of which have been directed towards mechanical conveying. Classification tests for characterizing material, specifically for pneumatic conveyor design applications, are still in the development stage and have not yet been published as a universally accepted standard. The publication by the Engineering Equipment Users Association is often cited (3).

Information on the behavior of bulk materials in conveying equipment is being developed through a number of sources. Extensive research on the characterization of bulk solids and the flow behavior of bulk solids in conveyors and feeders has been published (4). This work includes the design of devices to study and measure the abrasive wear of materials on surfaces that simulate conveyor components using the surface speeds normally encountered in mechanical conveying systems, and abrasive wear on surfaces that simulate silo walls at surface speeds normally encountered during flow. A device to measure the particle attrition that occurs during simulated flow or conveying is also described. Several proprietary devices have been developed from similar work (5). Useful information on the flow behavior of bulk powders and granules can be found in References 6–8.

Conveyor Types

Belt Conveyors. A belt conveyor is made up of an endless fabric or elastomer covered belt that traverses between two or more pulleys, and is supported at intermediate points by idler rolls. These conveyors can handle a wide range of materials, from fine powders to large, lumpy stone and coal. Material can be transported at rates of over 5000 t h and the conveyors operated at belt speeds ranging from 20 to 300 m min over very long distances. Versatility, reliability, and range of capacities have made belt conveyors the most commonly used bulk handling conveyors in industry.

A typical belt conveyor arrangement is shown in Figure 1. These conveyors can be arranged horizontally and with inclined or declined sections combined with convex and concave curves. The desired path of travel is limited only by the strength of the belt and the permissible angle of incline or decline for the particular situation.

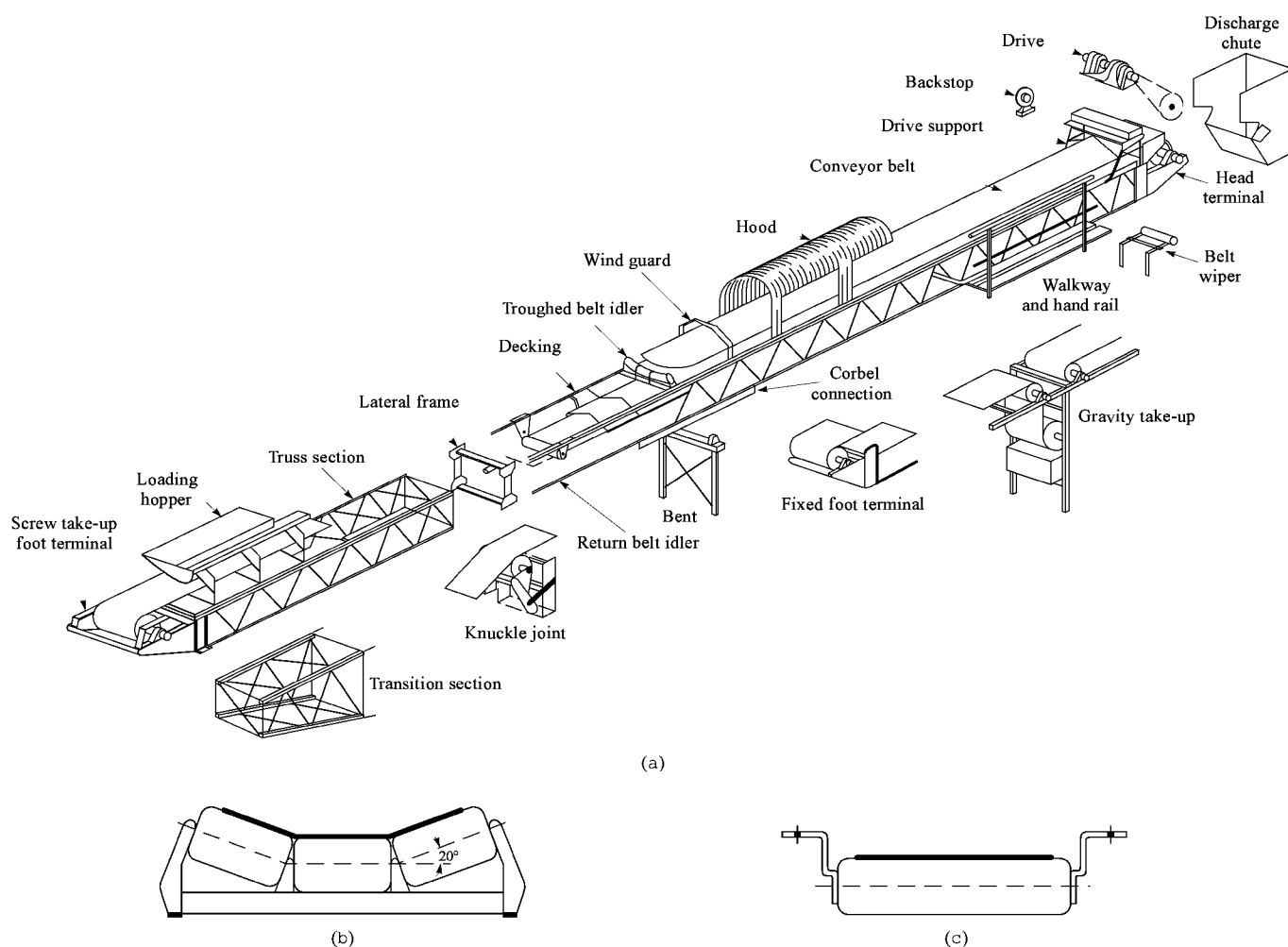


Fig. 1. (a) Belt conveyor and support assembly, (b) 20° troughing idler, and (c) return idler.

Bulk materials are sometimes conveyed on flat belts that are supported on horizontal idlers on the carrying and return runs. However, in most industrial systems, in order to increase handling capacity, the conveyor belt is formed into a trough shape after it has been loaded with material, and it is supported along the carrying run by troughing idlers. The most commonly used troughing idler consists of three rolls as shown in Figure 1b. An extensive guide to design and application of belt conveyors is available (9) as are belt conveyor standards (10–12).

Belt-conveying design technology is changing rapidly. Newer methods using computers for dynamic analysis can take the viscoelastic properties of the conveyor belt and the distribution of stresses during startup and shut down into consideration. Analysis is being applied to long conveyor belts and belts with horizontal curves. A description of these design techniques is available (13).

Material Characteristics Influencing Design. Materials ranging from fine powders to large lumpy materials, nonabrasive to very abrasive, free-flowing to cohesive, and nonfriable to friable can be handled on properly designed belt conveyors. Very sticky materials can be a problem, however, if these cannot be continuously cleaned from the belt surface.

Material characteristics typically used as criteria for determining the required belt width, the carrying capacity of a particular belt, and the maximum inclination at which the belt can be operated are (1) The angle of repose and the angle of surcharge. The angle of repose is the angle which a freely formed heap of bulk material makes with the horizontal. The measured value of this angle, when taken together with observations of particle shape and size, moisture and general flowability, determines what is known in belt conveying technology as the angle of surcharge. This is defined as the angle (to the horizontal) which a material assumes while at rest on a moving conveyor belt. For most materials, the angle of surcharge is usually 5 to 15° less than the angle of repose, but with some very free-flowing materials, it may be 20° less. The angle of surcharge is a basic design parameter for determining the transport capacity of a belt conveyor. Because the cross-sectional area of the heap on the belt is defined by the belt trough geometry and angle of surcharge, the belt speed required for a desired transport rate of a material with a known bulk density can be easily determined. (2) The size and proportion of lumps. The larger the lumps and the greater the number, the wider the belt must be to prevent the lumps from spilling over the edge of a horizontal conveyor, and the more likely they are to roll back, or fall over the edge, on an inclined conveyor. (3) Fluidizing or air retention properties. Materials that are easily aerated and have long air retention times impose limitations on allowable belt inclinations and belt speeds.

Carrying Idlers. The most commonly used troughed carrying idlers have the outer rolls inclined at either 20, 35, or 45° from the horizontal (see Fig. 1b). Carrying capacity of the belt increases as the idler angle increases. In the past, construction of multi-ply belts were fairly stiff, and did not allow the belt to conform well to idlers having troughing angles over 20°. The newer belts, having carcasses of various synthetic fiber blends, are more flexible and can be designed to trough at the higher angles (see FIBERS). As a result, the 35 and 45° idlers are gaining in popularity. Troughed idlers can be mounted on pivoted supports and spaced at intervals along the conveyor for training or guiding the belt, i.e., to keep it centered on the carrying idlers during transient upsets.

The mechanical design of the idler rolls is a function of the particular service under which the conveyor operates. Minimum industrial standards for roll dimensions, bearings, and application criteria for different service conditions have been established (14). Idler life is determined by a combination of factors such as bearings, seals, shell thickness, load density, and the operating environment.

Bearing rating, or bearing life, is the only variable for which laboratory tests can provide standard values. Thus CEMA uses bearings as a guide for establishing idler ratings (9). The term useful life (BU), representing the statistical point in hours where a minimum of 90% of the bearings are expected to be still functional with no increase in torque or noise, is employed. The minimum required load ratings for equal-length roll idlers for service conditions ranging from light duty to heavy duty, based on 90,000 h minimum bearing life at 500 rpm, have been determined for belt widths ranging from 450 to 2500 mm, for 20, 35 and 45° troughing angles, and for flat idlers. These ratings form the basis for the minimum design requirements for CEMA-rated idlers. Actual figures for idlers furnished by manufacturers must be obtained from the idler manufacturer. Whereas bearing life is useful as an indicator of idler life, other factors such as bearing seal effectiveness, may be more important in determining idler life under certain circumstances.

A variety of specialized idlers are available. Examples are plastic disk catenary idlers for handling wet corrosive materials; two roll idlers, where the rolls are oriented in a vee for lighter duty conveying system; and suspended idler supports for severe service. In this last type, three to five idler rolls are linked end-to-end and suspended from conveyor frame stringers to form a catenary that cradles the belt.

Conveyor belts are manufactured in widths up to 2500 mm. The belt represents a substantial part of the initial cost of a belt conveying system and it is the component most susceptible to damage. The typical belt consists of two principal elements: the covers (top and bottom), and the carcass. The primary purpose of the covers is to protect the carcass against environmental effects, wear, and cutting. Covers are natural or synthetic rubber, thermosetting elastomers, and thermoplastic materials. The carcass provides the tensile strength required to start and move the loaded belt, the transverse and longitudinal flexibility needed to allow the belt to both support the load and conform to the shape of the idlers when running empty and to properly wrap around pulleys, and the strength to resist impact forces.

Carcass and cover ratings for belts manufactured in the U.S. have been established by the Rubber Manufacturers Association (RMA) (15). Guidance for specific applications can be found in Reference 9. Similar, but not the same belt rating systems have been established in most other countries.

Carcass Construction. Carcasses are made of one or more plies of a woven fabric bonded together with an elastomeric compound. Woven materials that are used include cotton, rayon, nylon, polyester, aramids, and glass, in the pure form or in blends. The fabrics are constructed with warp yarns that run lengthwise along the belt, and filling (weft) yarns that run crosswise. There are a variety of fabric weaves available for specific applications (15).

Multiple ply belts manufactured in the U.S. were standardized on the basis of tension ratings into a range of MP designations that classified a belt tension rating in pounds per inch per ply, without regard for the type of fiber used. The development of the newer high tenacity synthetic fibers and improvements in belt construction led to a wide variety of belts having load capacities that equal or exceed those of the belts having the MP designation, and do it with fewer plies. As a result, the MP designation is less used. Belt tension ratings are being specified by the manufacturers on the basis of pounds per inch of belt width (PWT) for each particular belt of manufacture.

A more recent development in belt technology has been the solid-woven carcass belts impregnated and covered with poly(vinyl chloride) (PVC) or urethane plastisol. Steel cable belts, made with a single layer of parallel steel cables completely imbedded in rubber or enclosed between one or more fabric plies are used for very high tensile loadings that are beyond the capacity of the fabric-carcass belts. A more recent development in high tensile strength belts has been the use of Kevlar (DuPont), high strength aramid fibers in woven, cord, solid woven, or cable construction. A comprehensive survey of carcass constructions used in conveyor belt technology is available (16).

Take-Ups. A take-up is required on a belt conveyor to ensure the proper belt tension at the drive pulley and along the conveyor, as well as to ensure the proper troughing contour between idlers. A take-up is also needed to compensate for changes in belt length caused by elastic stretch during start-up, and any elongation characteristics of the belt that occur over a period of time.

Manually adjusted screw or ratchet take-ups that adjust the position of the tail pulley to control belt tension can be used on relatively short, light duty conveyors. Automatic take-ups are used on conveyors over about 25 to 30 m long. The most common is the weighted automatic gravity take-up (see Fig. 1a). Other types of automatic take-ups have hydraulic or pneumatic powered devices to adjust a snub pulley position and maintain a constant belt tension. The required take-up movement varies according to the characteristics of the belt construction and the belt length. Typically, take-up movements for plied belts are 2° to 3° of the center distance between head and tail pulley, and about 0.5° for steel cable belts. The take-up movements required for solid woven belts are usually shorter because of the lower elastic stretch. Take-up requirements for a particular situation should be confirmed by the belt manufacturer.

Backstops. A backstop is a device that permits rotation of the pulley in the forward direction but automatically prevents rotation in the opposite direction. A backstop should be installed at the headshaft of an inclined belt to prevent the belt from moving in reverse if the power to the motor is interrupted or if there is a failure in the mechanical drive system.

Belt Cleaning. Idlers and snub pulleys on the return run support the belt on the material carrying surface of the belt and, therefore, are exposed to any material that may cling to the surface of the belt. This material is then transferred to the surface of the return idler rolls and snub pulleys and can adversely affect the training and control of the belt path. Cleaning sticky materials from the surface of a belt is difficult. There are available a variety of single and multiple-blade belt scrapers having spring or counterweighted supports, motor driven or belt powered blade, or brush cleaners. Each has had varying success in thoroughly cleaning the belt surface. A variety of self-cleaning idlers constructed using rubber-disk, spiral, and beater rolls for use on the belt return run have been moderately successful in dislodging material from the belt surface.

Power. The power required to drive a belt conveyor is derived from the tensile forces required to propel or restrain the belt at the design speed. These include the tensile forces produced by the frictional resistance of the drive, conveyor components, and material; the acceleration of the material; and the gravitational forces required to lift or lower the material. Detailed information and methods of calculation can be found in belt conveyor design handbooks and in Reference 9.

Newer Designs. A number of belt conveyor designs that depart from troughed belt technology have come onto the market since the mid-1980s. These are expected to have a significant impact on belt conveying technology.

Flexible Sidewall Belt Conveyors. The flexible sidewall belt consists of three elements: a flat cross-rigid base belt, accordion-pleated flexible side walls to restrain the material being conveyed, and a cleat or divider spaced at intervals along the belt to prevent slideback of the material when the belt travels on an incline as shown in Figure 2a. The flexible sidewall allows the belt to pass around head and tail pulleys and bends on the vertical plane, so that it may be loaded and discharged on horizontal runs. The bend from horizontal to the elevating slope angle is achieved by means of deflection wheels that control the upturn of the belt. Sidewall heights range from 40 to 400 mm. Conveyor slope angles up to 90°, lift heights of 80 m, and capacities up to 10,000 t/h have been attained using these conveyors (16, 17). Another version of the flexible wall belt is the Camflex belt (Fig. 2b) which combines a flexible side wall with knobs projecting from the belt surface to provide frictional drag on the material to permit conveying at steep inclines.

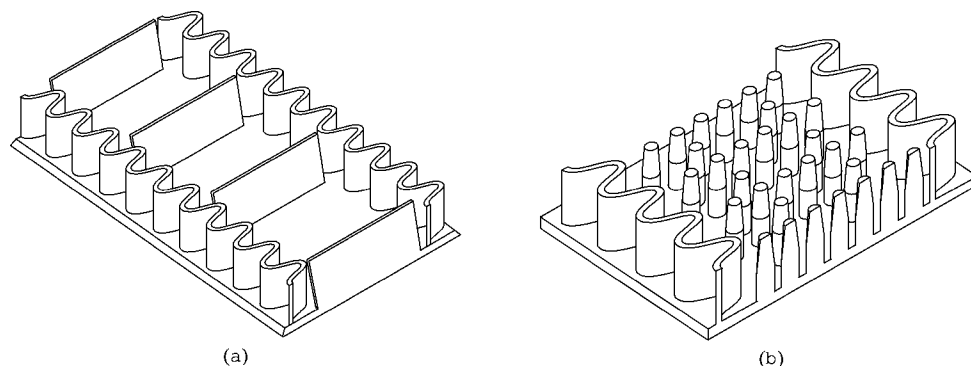


Fig. 2. (a) Flexible-wall belt (16,17); (b) Cambelt.

Serpentine Three-Dimensional Continuous Path Conveyor. The Serpentine conveyor (18), consists of a combination of conveyor track and belt guidance system. Convolutional belt pan modules are sequentially fastened together to form a continuous belting surface. The convolutions, which enable the belting surface to follow horizontal, vertical, and helical paths, also act as a cleat, permitting transport of loads at steep inclines. The convolutional belt is opened and flattened at the discharge, thus enabling cohesive or adhesive materials to be scraped from the belting surface. The modules are supported on a strand (or strands) of guided roller chain or drag chain, which provides the pulling element for moving the conveyor in a continuous path. These conveyors have been used for handling a wide variety of wet and dry bulk materials and industrial waste since the early 1980s.

Air Cushion Conveyors. Figure 3a shows an air cushion-supported belt conveyor. The belt and material are supported on an air film created by passing air through small holes or slots in a U-shaped trough beneath the belt. The air film reduces the conveying friction losses, resulting in a reduction in required

power and in belt wear as compared to idler supported belts. Belt widths range up 900 mm and transport distances up to about 90 m.

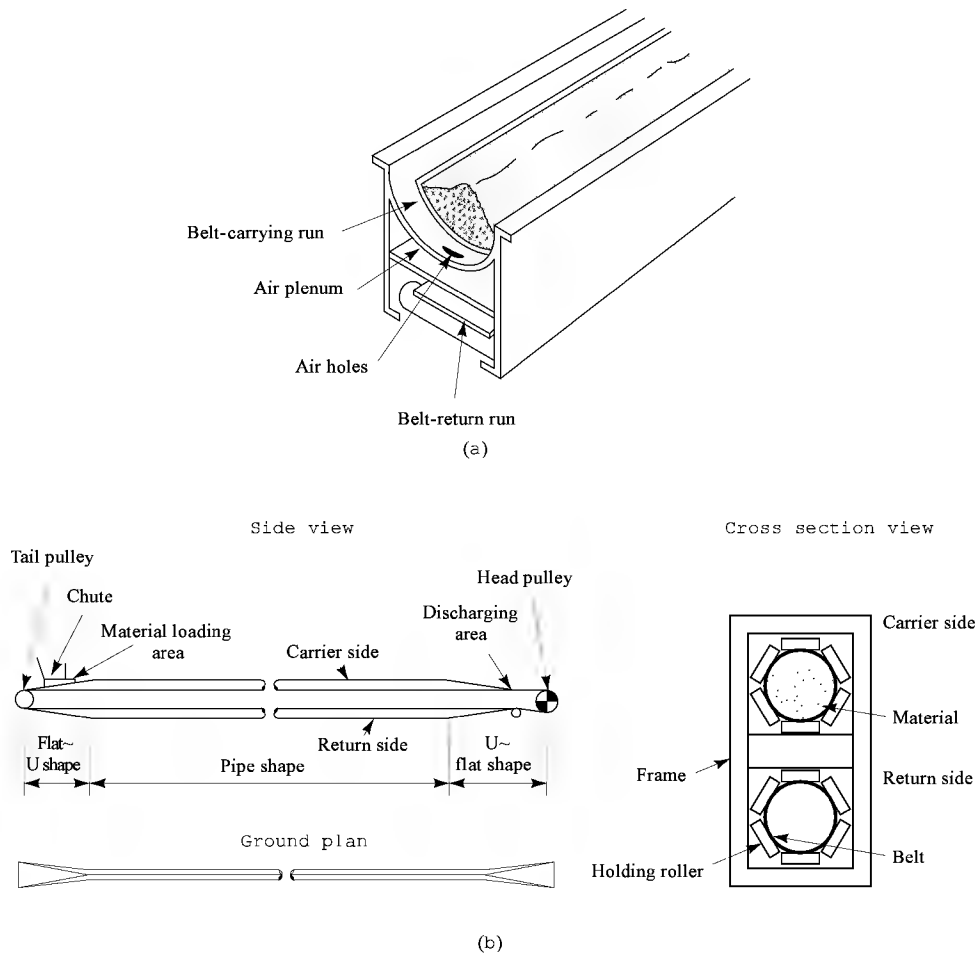


Fig. 3. (a) Air cushion-supported belt conveyor; (b) the Japan Pipe Conveyor.

Sandwich Belt Conveyors. The sandwich belt conveyor employs two rubber belts that sandwich the conveyed material between them, enabling the conveyor to transport material at inclines up to 90° from the horizontal. These conveyors have been operating in self-unloading bulk cargo carriers on the Great Lakes since the 1960s (19). They are now being designed and put in service for land-based conveying. There are two general types of designs: one uses a cover belt over a troughed belt (20), and the other two flat belts with the edges pinched together to restrain the material while moving it in the inclined or vertical lift section. Material is loaded on to these belts at a horizontal run before the incline, and discharged after the incline. Belt widths range from 900 mm to 1380 mm. The sandwich belts have been used for conveying at inclines ranging from 30 to 90°, at rates up to 400 t h.

Enclosed Tubular-Type Belt Conveyors. Manufacturers have been working to develop a belt that can totally enclose the material to be conveyed. This would minimize spillage and dust release into the environment, and would eliminate the need for cleaning the belt surface. Several designs have reached the commercial stage and have been installed in industrial plants. A review is available (21).

In the pipe conveyor, patented by the Japan Pipe Conveyor Company, the belt is in a flat position as it passes over the tail pulley to be loaded. It then passes through a transition section where it is transformed into a tubular shape with the belt edges overlapping to form a seal. The belt is constrained in this position by special multiroll idlers, spaced at intervals along the length of the conveyor. The belt is opened at the head pulley to discharge the material, and then reformed into the tube-shape as it continues back on the return run. The tubular belt configuration encloses the material, and thereby eliminates the need for weather and dust enclosures used on conventional belt conveyors. A further advantage is that the material handling surface of the belt never touches the idler surfaces, thereby eliminating the problem of return idler contamination. The diameter of the pipes formed by the belt range from 100 to 900 mm. An example is shown in Figure 3b.

The pipe conveyor can convey materials at higher angles of inclination than conventional trough belts and can be directed through very long radius curves in both vertical and horizontal planes. These features help to make it more competitive with conventional trough belts in those situations where a multiplane path using conventional belts would require using multiple belts and transfer points.

The belt is specially designed for this application. The compounding of the elastomer covers are controlled to reduce the pipe closing resistance and the plies in the carcass are stepped back at the edges to reduce edge stiffness and enhance the sealing when the belt is overlapped. The Rollgurt-Conveyor (22) is a similar type of enclosed tubular belt conveying system that was installed in an industrial plant in Sweden in 1990.

Folding Belt Designs. In folding belt conveyors, the belt edges are folded over the top of the material after it has been loaded on to the belt. The Goodyear folding conveyor belt carries material on conventional troughing idlers. The belt is specially constructed so the sides can be folded over the top of the material and constrained by guide rollers to enclose the material during transport over the troughing rollers. The U-Con belt conveyor (23) folds the belt and transports material over flat rollers.

Enclosed Pocket Belt Conveyor. The Sicon enclosed belt conveyor, shown in Figure 4, uses a special flat belt having flanged, cable-reinforced edges. After filling, the belt passes through a transition section where it is folded into the shape of a hanging pocket with both flanged edges pressed together by pinch rolls mounted at intervals on an overhead frame. The belt, having its edges constrained by the pinch rolls, and the material enclosed in the pocket, are transported along a path of guide and support rollers on an overhead frame. The path can include vertical and horizontal turns. The belt is opened at the head pulley to discharge the material and then reformed into the pocket shape for the return run.

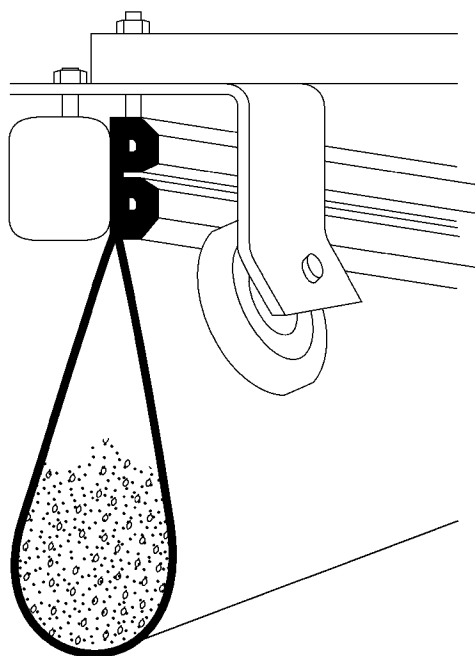


Fig. 4. The Sicon belt conveyor (22).

Screw Conveyors. A screw conveyor consists of a helical flight fastened around a pipe or solid shaft, mounted within a tubular-shaped or U-shaped trough. As the screw rotates, material heaps up in front of the advancing flight and is pushed through the trough. Particles in the heap, adjacent to the flight surface, are carried part way up the flight surface and then cascade down on the forward-moving side of the heap. Screw conveyors are of simple, relatively low cost construction, and are comprised of highly standardized component parts. These conveyors can handle a wide variety of solid particles ranging from lumps to powders within a completely enclosed housing at temperatures up to 275°C or higher if they are liquid cooled. The flights can be configured for particle mixing during transport, and flights and housing configured for particle cooling. Lumpy, sticky, or fibrous materials may cause problems in a screw conveyor. Conveying distances are limited by the torque capacity of available drive shafts. Power requirements are relatively high and conveying efficiency is considerably reduced when screws are inclined or mounted vertically.

Construction. The parts, dimensions, and dimensional tolerances used in manufacturing of screw conveyors are highly standardized. Standards for dimensions and minimum service requirements are available (24–26). These are accepted by most of the industry.

In a typical screw assembly, the flights are fabricated, then welded to a pipe that has bushings press-fitted or welded into each end to provide reinforcing for the conveyor couplings. There are two types of flights: helicoid and sectional.

Helicoid flights are formed by rolling a strip of metal so that one edge is compressed to half its original thickness, causing the strip to form a continuous helicoid having a uniform pitch. When this flight is mounted on the pipe shaft, the thinner edge is at the outside circumference. Helicoid flight conveyors are commonly fabricated in diameters up to 460 mm.

Sectional flights are formed by cutting a ring from a plate, making a radial cut in the ring, and cold pressing it in a die to form a single flight pitch (25). The flight thickness is the same at the outer and inner edges. The individual pitches are slipped onto the pipe shaft and butt welded to each other and to the pipe shaft. Sectional flight conveyors are commonly manufactured in diameters up to 600 mm. A sectional flight, which is more costly than a helicoid one, is used where the thicker flight edge can be used to advantage: it provides a longer service life when handling abrasive materials and additional strength to resist stresses caused when handling lumpy or compacting materials. Because each flight is made individually, the sectional flight screw is also used where varying or special pitch screws are required.

Screw conveyors are assembled by connecting screw sections in sequence within a trough enclosure, using screwed, flanged, or bolted connections. The most commonly used is the bolted connection using two special bolts spaced 90° apart, and inserted through the pipe, collar, and coupling shaft. The screw sections are supported at intervals within the trough by bushed hangers. The bushings, installed around the exposed portion of the coupling shafts between adjacent screw sections, are supported in the hangers that are fastened to the trough. Standard hanger spacing is from 3.45 to 4 m to match the standard screw lengths. The hanger bushings are exposed to abrasive wear from the material being transported, and must be replaced periodically. Hangers and bushings are an obstruction to flow and bushings are subject to wear; thus there are advantages to minimizing or eliminating hangers. This is often done by increasing the spacing between hangers, and using over-sized, larger diameter shafts to resist the increased bending stress. A typical screw conveyor assembly is shown in Figure 5.

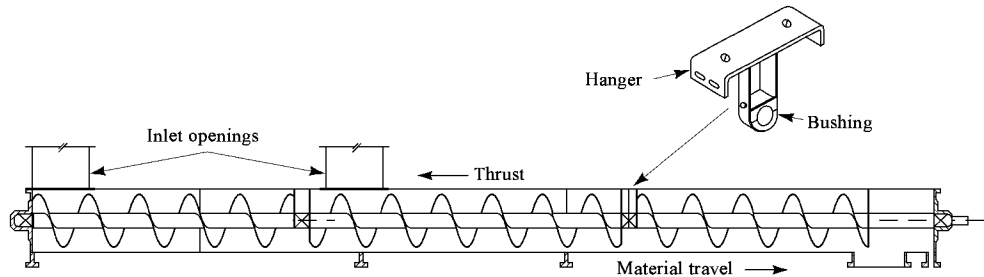


Fig. 5. A screw conveyor assembly.

Considerations for Sizing.

Lump Size. The amount, size, and character of lumps in the material to be handled is the first consideration when selecting the proper diameter of a screw for a particular application. If the lumps are expected to be friable, and easily broken in passing through the conveyor, or if the material contains no lumps, there is no limitation on the diameter of the conveyor, and selection can be made on conveying capacity alone. If the lumps are hard and not liable to be broken up during transit through the conveyor, then the screw diameter must have enough clearance between the pipe shaft and trough to accommodate these lumps. Manufacturers usually select the minimum screw diameter on the basis of maximum lump size and the percentage of those lumps in the material to be handled. CEMA recommendations (24) regarding acceptable percentages of lumps in mixtures of particles to be handled in screw conveyors is followed by most manufacturers.

Material Characteristics. In general, screw conveyors can handle all free-flowing materials. The more free-flowing, the lower the energy required to transport

the material. Cohesive materials and materials that tend to pack in the radial clearance between the flight tip and the trough require special considerations. These could include such things as closer than standard flight to trough clearances, tapered or wear-resistant flight tip surfaces, special nonsticking surfaces or materials on the flight, special flight geometries, and larger drives. The higher the filling (loading) in a trough, the more contact there is between material and the hanger and bushing. To assure a reasonable service life for the bushings and the coupling shafts that run in these bushings, the conveying industry has established a guide for determining the maximum depth of material that should be carried in a screw trough for particular materials.

Standard practice calls for limiting the depth of loading in a trough where hangers are used to 45° of the cross-sectional flow area when handling very free-flowing fine or granular, nonabrasive materials, 30° for less free-flowing materials and for moderately abrasive materials, and 15° for very abrasive materials. The reason for the limitation on cross-sectional loading is to avoid jamming material around the hanger and to achieve a reasonable service life for the bushings and flights. However, if there are no hangers in the conveyor, trough loadings up to 90° are possible. The loose bulk density for a material is used for calculating actual cross-sectional loading in a screw conveyor. The actual bulk density of material as it is exposed to agitation as it moves through a screw conveyor can only be estimated from loose densities determined by bench-top tests.

Flight Geometry. The flight pitch effects screw conveying efficiency. A pitch-to-diameter ratio of one has been found to provide the best combination of efficiency and cost effectiveness for horizontal screw conveyors. Standard screws, therefore, have a pitch equal to the diameter. There are a number of other flight configurations used for special purposes that include feeding, mixing, gas stripping, and preventing flushing. A description of these, and others, can be found in the literature (24). Information on screw conveyors is also available (27–29)

Screw Inclination. The angle at which the screw is inclined to the horizontal affects conveying efficiency. As a screw is inclined, the slope of the flight surface becomes closer to horizontal and becomes less effective in moving material forward. At some point, depending on the material properties and the conveyor geometry, the increased turbulence and tumbling of the material causes a portion of the material to spill back over the pipe and over the flight tips into the preceding pitches. This reduces the conveying efficiency and increases the cross-sectional loading. Higher speeds and additional power are then required to achieve capacities comparable to a horizontal screw. As a general rule when the angle of inclination exceeds 10 to 15°, modified designs are required in order to offset the reduction in conveying efficiency. These modifications can include close clearances between the flight and trough, a shorter screw pitch, a tubular housing in place of a U-trough, and increased screw speed. Information on design of inclined screws can be found in Reference 24.

Vertical Screw Conveyors. Many free-flowing materials can be conveyed vertically with a screw. Screw elevators, designed for this purpose, have tubular housings, run at speeds ranging from 200 to 400 rpm, and have volumetric efficiencies of about 25° of an equivalent horizontal screw. Most of these machines are limited to vertical lifts of about 9 m, although some units have been built for lifts up to 30 m for handling grain and other agricultural materials. The vertical screws are not self-cleaning; ie, if material feed to the inlet at the bottom stops, even though the vertical screw is rotating, some material remains at the bottom until more material is fed into it.

Screw Conveyor Capacity. The volumetric capacity of a horizontal screw conveyor is calculated on the assumption that all material contained within one screw pitch moves one pitch distance in one screw revolution. Volumetric conveying capacity is calculated as

$$\text{volumetric capacity} = (A_s - A_p) P \cdot K \cdot N$$

where A_s is the cross-sectional area of the screw flight, A_p is the cross-sectional area of the pipe shaft, P is the flight pitch, N , the rotational speed, and K is the percentage of the cross-sectional annular space between flight and pipe shaft occupied by the material. $K = 0.45$ for fine to granular, free-flowing, nonabrasive to mildly abrasive particles; $K = 0.30$ for fine to small lumps, average flowability, mildly abrasive particles; and $K = 0.15$ for very abrasive materials. This procedure ignores the volume occupied by the flight itself, but is sufficiently accurate for most engineering purposes.

Power to Operate a Screw Conveyor. The power required to operate a screw conveyor is dependent, to a large extent, on the handling characteristics of the material to be transported. Formulas for calculating power use empirically derived factors to account for the conveying characteristics of specific materials, the configuration of the screw, and the bearing friction. These formulas have been developed by CEMA and can be found in the literature (24,25) and in engineering handbooks. It is assumed that the total power is equal to the sum of the power required to overcome friction and the power required to transport the material.

Bucket Elevators. In a bucket elevator, a series of buckets attached to an endless belt or chain are filled with material and lifted vertically to a head pulley or sprocket, where the material is dumped. The buckets are then returned back down to a tail pulley or sprocket at the bottom. Bucket elevators are not self-feeding. They must be fed at a controlled rate to avoid overfilling the buckets and damaging the machinery. In the usual arrangement of a bucket elevator, the chain or belt path is vertical or steeply inclined in a single plane. Special chain supported bucket systems that can travel in two and three planes have been developed.

There are four broad classifications of bucket elevators: centrifugal, continuous, positive, and internal discharge. Centrifugal and continuous discharge elevators are by far the most commonly used. Two specialized versions used for high capacity handling are the super-capacity elevator and the cement mill elevator. The positive discharge and the internal discharge elevators are used for special applications. The various elevator types are shown in Figure 6.

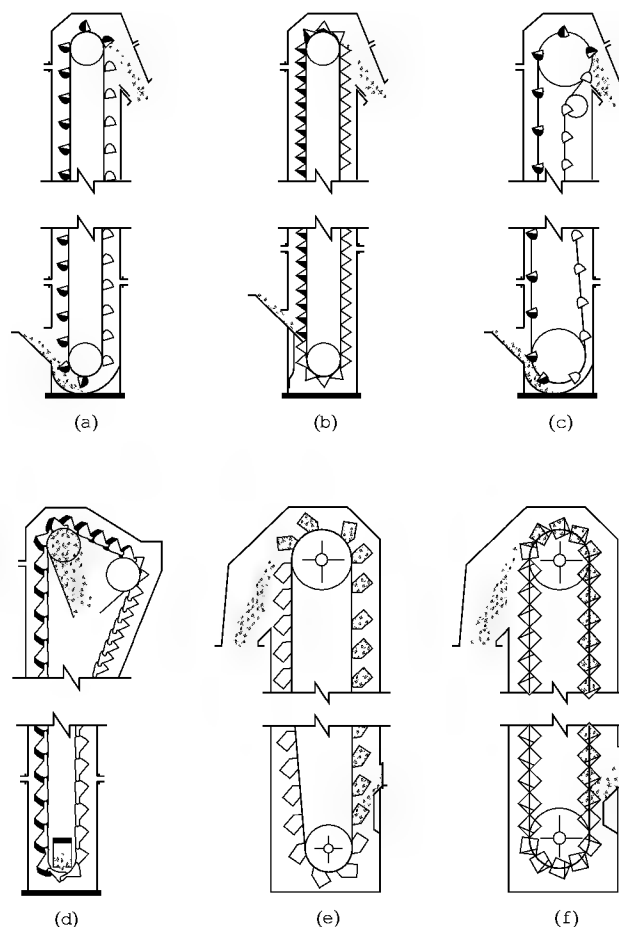


Fig. 6. Bucket elevators: (a) centrifugal; (b) continuous; (c) positive; (d) internal discharge; (e) super capacity; and (f) cement mill.

Centrifugal Discharge. In centrifugal discharge elevators, the buckets are spaced apart on the chain or belt. Material is scooped from the boot and then discharged by centrifugal force as the buckets approach and pass over the head pulley or sprocket. Speed of the chain or belt is critical to proper discharge of the material. The critical speed has been defined as the speed at the point where the centrifugal force at the center of mass of the material in the bucket is equal to the gravitational force (30). The best operating point for an elevator is where the centrifugal force is two-thirds the gravitational force. The effects of bucket geometry and speed on material discharge have also been reported (31,32). Changes in exit trajectory of the material leaving the bucket as a result of material friction on the bucket wall and the change in center of gravity of the mass as it slides along the wall have been investigated.

Industrial, centrifugal elevators usually operate at speeds of about 75 m/min, and handle free-flowing, fine and loose materials having lump sizes of <50 mm. Sticky material can be a problem. Fine fluidizing materials often require perforations in the bottom of the buckets to vent entrapped air. Centrifugal elevator capacities range up to 370 m³/h for a single row of buckets, and up to 1400 m³/h for multiple rows of buckets. The buckets can be mounted on a belt or chain.

Continuous Discharge. Buckets are mounted end-to-end on the chain or belt in continuous discharge elevators forming a continuous strand. Material is fed through an inlet chute into the rising buckets as they pass through a loading leg. The loading leg is formed of fixed vertical plates mounted within the boot, in close proximity to each side of the rising buckets, to confine and direct the material stream. The material, which discharges by gravity as the bucket passes over the head pulley or sprocket, slides over the inner surface of the bucket, out over the back of the bucket immediately ahead, where the trajectory carries it out through the elevator discharge spout. The back of each bucket is fashioned with projecting sides that form a chute to contain the flowing material from the preceding bucket.

Continuous elevators can handle a wide range of materials from light to heavy, as well as free-flowing granular and pulverized materials containing lumps up to 100 mm. This elevator handles material more gently than a centrifugal elevator. Bucket speeds range from 30 to 46 m/min. The lower speeds are used in order to properly fill the buckets when fluffy, low density materials are handled. Capacities range up to 75 m³/h on standard elevators.

Super-capacity elevators are continuous type elevators in which the buckets are mounted between two strands of chains. This arrangement makes it possible to handle a larger volume because the bucket can be extended in back of the chain. Elevators of this type are capable of handling up to 375 m³/h.

Cement mill elevators, also a version of continuous elevators, use a pocket-shaped bucket (Style AC in the U.S.), mounted on a single strand of chain. The buckets can be closely spaced to pick up material fed through a loading leg, or can be more widely spaced to pick up material from the loading leg, as well as scooping from the boot. Although originally developed for cement handling, this elevator is also used for handling very fine, aeratable powders. The bottom of the buckets have air release holes to allow entrained air to escape during filling, thereby increasing filling efficiency. Capacities range up to 210 m³/h.

Positive Discharge. In positive discharge elevators, spaced buckets are mounted between two strands of chain and the buckets are discharged by snubbing the chain after the head sprocket, so that the buckets are inverted over the discharge spout. These elevators are used for handling materials that tend to stick in the buckets or for low density materials that do not discharge readily. Capacities range up to 40 m³/h.

Internal Discharge. Continuously overlapping buckets supported by a strand of chain at each end are inverted as they pass over the head sprockets in internal discharge elevators. Material is loaded into the bucket through a chute extending into the side of the elevator boot section, and is discharged through a chute extending out the side of the head section. This design is useful for gentle handling of small particles such as plastic pellets, granular chemicals, agricultural products such as seeds, shelled nuts, etc., and even mechanical parts. Capacities range up to 23 m³/h.

Buckets. Elevator capacity is a function of bucket volume as well as speed and spacing on the chain or belt. There are a variety of bucket geometries, and selection is based on the materials to be handled. Buckets used in centrifugal discharge elevators are shaped and reinforced so that material may be easily scooped or dug from the boot at the lower end of the elevator where the material enters. Buckets used in continuous elevators are shaped to receive material that flows into them from the inlet chute in the boot. Most of the buckets can be furnished in cast malleable iron, carbon, galvanized and stainless steel, and other corrosion-resistant material. Buckets are also available in cast and injected molded nylon or polyethylene. The polymer buckets weigh substantially less than the metal ones, so the chain or belt loading is reduced when used. These synthetic buckets can be used for fine abrasive materials and in many cases slightly adhesive materials that do not readily release from a metal bucket.

Casing Enclosure. Steel casings are normally furnished in intermittently welded, dust-tight construction which is the lowest cost construction. Weather-tight, watertight, or gastight can be specified. Casings can be fabricated of metal or fiberglass. Except for very large units, bucket elevator casings are self-supporting against vertical loads imposed by the buckets, chain, material, and drive. However, the casing must be braced or anchored at appropriate intervals to maintain stability.

Chain. Chain, which can be used in all types of elevators, must be used for handling hot (>120° C) materials. Several types of chains are commonly used. Selection depends in large part on the abrasiveness and temperature of the material and the required working chain loading (chain pull) for the particular application. The three types of chains most commonly used in industrial elevators are (1) steel side bar, bushed chain, or steel knuckle chain made entirely of alloy steel. Hardened steel bushings with flat surfaces on each end are press fitted and locked into similar shaped holes in the side bars, forming a single link. The links are connected together by side bars and hardened pins that pass through the bushings and are locked to prevent rotation. These chains are used for heavy-duty elevator service. (2) Welded steel chain which is dimensionally similar to the knuckle chain. Hardened steel barrels are welded between steel side bars and the links are connected by heat treated pins. (3) Combination chain is made up of cast malleable iron links alternating with steel side bars. The cast links are joined by hardened steel pins inserted through machined holes in the steel side bar and cored holes in the cast link. The pins are locked against rotation by a machined flat surface on the end of the pin engaging a similar shaped hole in the steel side bar. These chains are commonly used for less severe elevator duty.

Belting. In a belt elevator, the buckets are fastened to the belt with special flat-headed bolts. Belts used for bucket elevator service have the same type of construction as those used on belt conveyors but the selection criteria differs in one important respect: the belt must have sufficient body strength to prevent the bolt heads from pulling through the belt carcass. The carcass must resist the stress on the bolts caused by the digging of the bucket in the boot, and the prying forces caused by material lodging between the belt and the back face of the bucket as the belt bends around the pulleys. In many cases the bolt pull-out rating of the belt determines the selection of the belt construction rather than the working tensile loading. Resistance to bolt pull-out is provided by the appropriate number and thickness of plies as well as the ply material.

Belt tensile strength must withstand the working load that includes weight of the belt, buckets and material suspended on the carrying side, digging and filling loads, friction in the system, and the additional loads required to provide driving friction. Procedures for estimating the digging and friction loads, and bolt pull-out ratings of belts, can be found in elevator or belt manufacturers handbooks.

Elevator belts can be spliced endless by bolting overlapping ends (lap splice); butting the ends then overlapping them with a belt strap, and bolting them together (butt strap splice); connecting butted ends with a bolted hinged plate-type fastener (mechanical fastener); clamping butted ends (clamp fastener); or by overlapping the plies and vulcanizing them together. Lap and butt strap splices are most commonly used.

Head pulleys for belt elevators are crowned to help train and guide the belt and quite often are rubber lagged to increase traction with the belt. Wing-type pulleys, which resemble a squirrel cage, are often used as a tail pulley to minimize any pinching of materials between the belt and tail pulley surface.

Take-Up and Hold-Back. An elevator chain wears and elongates, and a belt stretches during service life. A chain also elongates when handling hot materials. Therefore, a take-up adjustment is needed to maintain tension between the head and foot shafts. A manually adjusted screw take-up that moves the tail shaft or head shaft or a self-adjusting weighted take-up that maintains a constant gravity force on the tail shaft may be used. The gravity take-up must be used when handling hot materials in order to maintain chain and sprocket tooth engagement, as the chain length changes with thermal expansion.

The weight of material in the buckets on the loaded side of an elevator chain causes the elevator to momentarily run backwards if, during operation, the power is interrupted or there is a failure in the driving system. Because this could be a hazard to operating personnel, as well as damage to the elevator, a backstop, similar to that described for a belt conveyor, should be used.

Vibrating Conveyors. A vibrating conveyor consists of a trough supported by tuned springs and or hinged links having a drive system. The drive system is arranged to oscillate the trough, causing solid particles to be moved along the trough. Thus these conveyors are sometimes called oscillating conveyors. There are two types of oscillating conveyors: reciprocating and vibrating.

On a reciprocating conveyor, material is carried forward in a horizontal direction by frictional contact with the trough. Inertia causes the material to be left in that position as the trough is quickly returned to the initial position. These conveyors are useful for handling granular free-flowing materials with a minimum of attrition.

On a vibrating conveyor, material moves along the trough in a series of hops. The particles are accelerated from the trough in an upward and forward trajectory as the trough moves forward. The particles return to the trough in the forward position as the trough completes its return stroke. The vibrating conveyor is the most commonly used type of oscillating conveyor because of its flexibility and ability to handle a wide range of materials.

Material must be fed onto a vibrating conveyor at a controlled rate. These conveyors are not designed to operate under a head load from solids in a storage silo or hopper. In addition to horizontal conveying, these conveyors can be used to perform other functions such as elevating, heating, drying, cooling, fluidizing, agglomeration, screening, and dewatering (qv).

Vibratory conveyors are capable of handling a wide range of materials. They provide gentle handling of food products such as friable flakes and pellets, pharmaceuticals, powdered and granular chemicals and minerals, and discrete metal parts. They are uniquely suited for handling abrasive, hot, and dusty materials and can be designed to withstand heavy impact loads from materials such as rocks, iron and steel castings, and metal and wood scrap. They operate at frequencies ranging from 5 to 15 Hz; strokes range from 50 to 5 mm; and lengths are up to 50 m.

Material Handling and Capacity Considerations. Dry granular material is almost always easily handled on a vibratory conveyor. Many fine powders and pulverized materials move more slowly on a vibrating trough and these tend to deaerate and form a relatively impermeable, slowly moving layer when exposed to the vibration. As a result, these types of materials can only be transported in relatively shallow layers on the trough. Wet, fine materials or adhesive materials do not move well on a vibrating trough. Methods used to improve flow along the trough include: heating the trough, lining the trough with an elastic polymeric material, or using a very long stroke on the pan oscillation.

Linear transport velocity of the material is almost proportional to the product of frequency and stroke for these conveyors. However, the transport velocity and the depth that can be obtained with a specific material is dependent on the handling characteristics of that material. Manufacturers have accumulated data banks relating material characteristics to attainable conveying velocities for their proprietary conveyor designs.

The movement of material on vibrating troughs has been studied and there has been some success in estimating the theoretical velocities of free-flowing granular materials in single array or in thin beds, using empirical material flow factors derived from tests (33). These correlations based on single particle or thin layer trajectories are not representative of real-life, thick bed situations. The complexities involved in predicting the influence of vibration on material, the handling properties of which cannot be well defined, makes it prudent to conduct tests to confirm the design parameters for any vibrating conveyor application.

Drive Systems. Positive, fixed displacement mechanical drives are the most commonly used. These conveyors consist of a trough and a spring system, usually fiber glass or metal, that supports and guides the trough from the rigid base, and a motor driven crank and connecting rod that drives the trough. The rotation of the crank arm produces a fixed trough displacement. The drive system is tuned to operate subresonant, ie, just below resonance or natural frequency with the springs and trough. Much of the energy is alternately stored and released back into the system by the springs. Thus the only power needed is that for starting and for making up for losses resulting from friction and air resistance.

Leaf springs are used for reaction and support on light to medium duty applications. For heavier duty applications, stabilizer links are used for supporting the trough and tuned coil or elastomer springs are used for reaction. Shock absorbing devices are used in the drive train to isolate shock loads developed during starting. The direct mounted rotating eccentric mass design makes use of a rotating exciter, consisting of adjustable-position eccentric weights mounted on the shafts of a double-shafted electric motor, which is mounted directly on the trough or on a counter-balance conveyor base. Because these machines do not have a positive crank arm drive, they do not have a fixed amplitude. This type of conveyor also operates at subresonance and is fine-tuned by weight adjustment on the trough before being put in service. This design requires fewer moving parts than the direct drive machine, but requires more attention to design and tuning. The exciters usually operate at frequencies of 15 to 20 Hz.

Excitation can also be supplied by an electromagnetic exciter that uses a rectified, pulsed a-c power supply, or a-c supply to an apposed electromagnet permanent magnet drive. These units operate at very short strokes and frequencies of 50 to 60 Hz. Although designed primarily as feeders, they are occasionally used as short conveyors.

Compressed air or hydraulically driven reciprocating piston or rotary exciters are sometimes used in short conveyors. They are particularly useful where explosion hazards limit the use of electrical drives.

Design Variations. Vibrating conveyor troughs can be custom-designed for particular needs. Troughs can be rectangular, with or without

covers, and they can be tubular. Tubular troughs are useful where dust-tight operation is required. Inclined flat bottom troughs with a saw-tooth-shaped carrying surface can be used for conveying granular materials up 10 to 20° slopes. Troughs can be provided with multiple discharge openings and can be designed with expansion decks to handle materials at temperatures up to 538°C. Tubular, or rectangular troughs, wound in a vertical spiral, can be used for vertical conveying of granular materials.

Isolation Mounting. The static and dynamic forces that are imposed on a supporting foundation or supporting structure by a vibrating conveyor are an important consideration in conveyor selection. Static forces are vertical forces resulting from the weight of the conveyor assembly and material in the trough. The dynamic forces act in the line of action of the conveyor and can be resolved into vertical and horizontal components. The structures supporting these conveyors must be rigidly designed to withstand cycling forces. The conveyors can be designed to minimize the forces transmitted to the supporting structure. This is particularly desirable when the conveyors are to be mounted on elevated supports, or on the upper floors of a building. Methods to reduce transmission of forces to supporting structures include: hanging the conveyor on flexible cables, mounting the conveyor on a tuned spring supported weighted base, or using a counterbalance mounted on a tuned spring system identical to that on the conveyor trough and driving it 180° out of phase with the trough. The dynamic forces transmitted into supporting structures have been reduced by up to 95% by these methods.

Information with regard to isolation mounting can be found in References 33–37.

En-Masse Conveyors. An en-masse conveyor consists of an endless chain or cable, pulling a series of spaced skeleton or solid plug flights through an enclosed casing or housing. Material is introduced through an opening in the casing, where it is captured by the flights and drawn through the casing until it reaches an opening in the housing, where it discharges by gravity.

En-masse conveyors offer unique advantages over other types of conveyors: they are compact in cross section and totally enclose the bulk material; they can be made vapor and gas tight; they can handle many materials with little particle attrition; they can have an L-shaped or Z-shaped path, thereby eliminating transfer points that are required by conventional straight line conveyors; they can combine feeding, conveying, and elevating in one machine; and they can have multiple inlet and discharge openings.

Material Characteristics. Performance of an en-masse conveyor is very dependent on the characteristics of the bulk materials handled. This conveyor is best suited for nonabrasive, free-flowing, granular or powder materials. Sticky or smearing materials can build up in the clearance between flight and casing causing a mechanical overload. Particles enter into these conveyors between the flights as they move across the feed opening in the casing. The chain or cable restricts the depth of the opening into which any lumps that may be present must fall. If a lump does not drop completely into the opening before the advancing flight carrying it reaches the enclosed portion of the casing it gets pinched between the flight and casing. The conveyor chain or cable is then exposed to a potentially damaging shock loading. For this reason, the potential maximum lump expected to be present in a material must be made known before the proper size conveyor can be selected. As a rule of thumb, the largest lump to pass into the conveyor should be no larger than one-half the opening between the tension member of the chain or cable and the casing. Very fine powders that aerate readily can also be a problem. It is prudent to test material in an en-masse conveyor in the manufacturer's laboratory if there is no prior experience with the specific material to be handled.

Chain En-Masse Conveyor. The oldest and most well-known en-masse conveyor for heavy-duty service is the skeleton-flighted Redler conveyor. The open skeleton flights are cast carbon steel or stainless steel. Flights are enclosed in rectangular casings; widths range from 127 to 590 mm. Volumetric capacities range from 0.03 to 17 m³ min and chain speeds from 1.55 to 60 m min. Typical speeds for some materials are: 9 m min (fly ash), 12 m min (coke), 9–18 m min (granular chemicals), 38 to 46 m min (grain) and 30–46 m min (wood chips). These conveyors can be arranged in a L, Z, and loop path, as well as in a horizontal loop or horizontal to inclined path. They can have multiple inlet and discharge points.

Other chain-type en-masse conveyors use flight configurations made up from plates or bars welded to standard forged chain and mounted in a rectangular enclosure.

Tubular En-Masse Conveyors. Lower cost tubular en-masse conveyors for light duty conveying are also available. There are two general types of construction: solid disks connected by chain links or disks connected by a steel cable. The tubular casings (or pipes) range from 50 to 300 mm diameter. Plugs (flights) are nylon or polyurethane and the casings are carbon steel or stainless steel. These units are able to handle sludge as well as dry powders. A specialized type of disk chain conveyor called the Aeromechanical conveyor operates at 220 m min to handle dry, free-flowing powders having cable-driven disks operating in 100 to 200 mm diameter tubes. The tubular conveyors can be arranged in paths similar to those of the Redler.

Other Chain-Type Conveyors. A number of driven chain conveyors are used for conveying bulk solids in horizontal or inclined paths. These conveyors are made up of standard mechanical components that can be configured in different ways to perform a particular task. The types most used in the chemical industry include apron, flight, and drag chain conveyors.

Apron conveyors consist of a series of overlapping or interlocking pans on which material is carried. The pans can have a variety of shapes to fit the required service. The sides are usually extended to contain the conveyed material. The pans are supported on bushed or antifriction bearing shafts and rollers, or flanged wheels, which operate on a track. These conveyors can handle materials over horizontal, inclined, or a combination of horizontal and inclined paths. They handle practically any bulk material including ores, gravel, coal, bulk chemicals, scrap, and refuse materials. An apron conveyor is ideally suited for handling heavy, lumpy and abrasive materials that would be unsuitable for a belt conveyor. Fine material dribbles between the overlapping pans. Apron widths range up to 1600 mm in width and conveyor speeds run up to 19 m min.

A flight conveyor consists of one or two endless power driven chains carrying spaced scrapers or flights for pushing material along a stationary trough. Discharge can be at the end of the trough or at intermediate points. These conveyors can be arranged horizontally or at an incline. They can handle granular, lumpy, abrasive materials, as well as nonabrasive powders, in concrete, steel, and wear-plate-lined or T-bar-lined, troughs. They can be designed to handle hot materials up to 1000°C using air or water cooled troughs or water submerged troughs.

A drag chain conveyor consists of a single strand of endless cast or welded steel chain that pushes or drags material through a trough. It is a simple, low cost conveyor for handling ash, coal, hot clinkers, and scrap waste materials. A variety of cast iron, cast steel, and welded steel chains are used for these conveyors.

Air Activated Gravity Conveyor. The air activated gravity conveyor is a very simple, inexpensive, maintenance-free, and low power-using device for conveying fine, easily fluidized, and aeratable powders. Originally developed as the Air Slide conveyor, it consists of a downward-sloped rectangular trough bisected by a porous membrane that defines a lower and upper channel in the trough. When air is supplied to the lower channel, it permeates through the membrane, and aerates powder resting in the upper channel, causing the powder to flow like a liquid down the surface of the inclined membrane. Powder flow occurs even if the membrane is only partially covered with material, because the pressure drop through the membrane is such that the air flow is uniformly distributed across the membrane surface independent of the depth of material above.

There are two types of these conveyors: closed and open. In the closed, the air and powder flow channels are enclosed within the trough. The air that permeates through the membrane exits with the powder in the upper channel. Enclosed types are used for transporting powders from point to point. Open-type air activated conveyors are mounted at the bottom of silos and hoppers to act as flow promotion devices. The upper (powder) flow channel is omitted. When air permeates through the membrane and the surrounding powder, the air causes the powder to flow towards the silo outlet. Multiple, open-type units can be arranged in a variety of configurations and they can be activated in a predetermined sequence to control the powder withdrawal pattern within the silo. Trough widths in commercial conveyors range from 100 mm to 850 mm. Nominal capacities for these widths range from 13 to 1500 m³ h.

The membrane is usually made from one of several materials. Woven polyester or cotton, the most commonly used and least expensive material, is adequate for temperatures up to 150°C. Sintered plastic is used where a low cost, washable surface is desired. This material is temperature limited by the polymer material to about 60°C and the flow of some powders may cause a static charge build-up on the membrane that could be hazardous in some operations. Woven fiberglass fabric or porous ceramic block is used for temperatures up to about 425°C. Sintered stainless steel powder or bonded stainless mesh is used for corrosion resistance, and for temperatures up to 530 to 650°C. Additional information can be found in the literature (38,39).

Pneumatic Conveyors. In a pneumatic conveyor, bulk solids particles are transported through a closed duct in a gas stream. A wide range of particle sizes, from powders to large chunks and from fibers to chopped sheet waste, can be handled through these systems. The only significant restrictions are that the conveying pipe should be able to accommodate the largest particles, and the material should not be sticky. Pneumatic conveyors have a number of important advantages over mechanical conveyors that have led to widespread use: they provide great flexibility and compactness in system arrangement; they can be arranged with a multiplicity of solids pick-up and discharge points; they can provide a complete enclosure for protection

from contamination of the operating areas, or contamination of the materials being handled; and they can be designed to provide heating, cooling, drying, or mixing during transport. Recirculating inert gases can be used for safely conveying explosive, toxic, or other sensitive materials.

Disadvantages are more power is required to operate a pneumatic conveyor, compared to a mechanical conveyor for the same capacity and transport path; they are limited to about 50 t/hr capacity in industrial nonagricultural applications; the conveying pipe, bends, and fittings are subject to erosive wear when handling abrasive materials, and friable materials may be damaged during conveying; the pneumatic conveying process is more sensitive to the characteristics of the material to be handled than are most mechanical conveyors. The handling characteristics must be known before a conveying system can be designed. If no prior experience with a particular material in question is available, testing of that material in a pneumatic conveying rig is mandatory in order to produce reliable design data.

Classifications. Pneumatic conveyors are commonly classified as being either a dilute phase, or a dense phase conveyor. The differences between the two can be illustrated by the use of a typical phase diagram (40) to characterize the flow conditions that occur during the conveying process. A typical phase diagram for horizontal conveying is shown in Figure 7, where the pressure drop per unit length of pipe, $\Delta P/L$, is plotted against the average superficial air velocity in the pipe. Line AB represents the pressure drop with air alone flowing ($G_0 = 0$). When solids are introduced into the conveying line at a constant rate, G_1 , the pressure at Point B₁ increases to point C₁, because of the increase in gas-particle drag. If the gas velocity is then reduced, without changing the solids mass feed rate, the solids velocity decreases, the solids loading (mass solids/mass air) therefore increases, and the unit pressure drop follows the path C–D. At point C, the particles are fairly dispersed in a suspension across the pipe cross section, but as the gas velocity is lowered to approach point D, the solid concentration in the moving stream begins to become more concentrated near the bottom of the pipe, and eventually takes on the appearance of a moving rope or strand. As the air velocity is further reduced, a point of minimum pressure, Point D, is eventually reached and at that point, particles begin to settle or salt out along the bottom of the pipe. The air velocity at that point is called the saltation velocity for that specific material in that specific pipe diameter. It has been reported (41) that powders begin to settle just before the minimum pressure point has been reached, whereas large granular particles begin to settle at the point of minimum pressure.

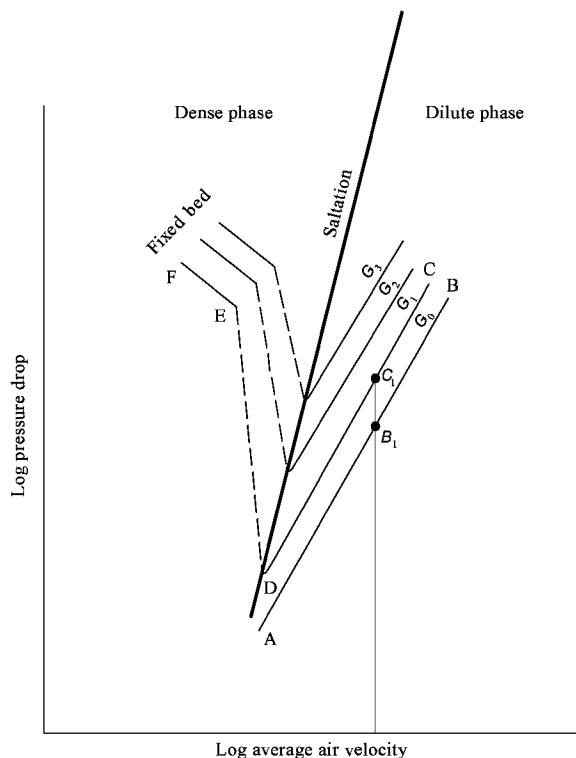


Fig. 7. Phase or state diagram for horizontal conveying where G_i represents a particular mass flow rate, line AB corresponds to the pressure drop for air alone flowing in the system, $G_0 = 0$, and (—) is the minimum pressure line where saturation occurs. Other points are explained in text.

If the pressure conditions at higher solids mass flow rates (G_2 , G_3) are superimposed on the phase diagram, the minimum gas velocity required to maintain conveying increases as the solids loading is increased. A line drawn through the minimum pressure points on the phase diagram defines the boundary conditions between what is commonly called dilute phase, to the right of the line, and dense phase conveying, to the left of the line. For a specific material the minimum velocity to maintain conveying increases as the pipe diameter is increased. An intermediate phase between dense and dilute has also been defined to be that condition existing at, or near, saltation conditions, where strands are observed. This definition has not been universally accepted.

A phase diagram for vertical conveying can be constructed in a manner similar to the horizontal phase diagram. The minimum superficial air velocity in a vertical pipe, analogous to the saltation velocity, is called the choking velocity. The minimum air velocity for avoiding choking in a vertical pipe is lower than that required to avoid saltation in horizontal pipes. Because conveyors are usually composed of horizontal and vertical sections, the minimum air velocities used for design of conveying systems are those velocities necessary to avoid saltation in the horizontal sections.

The phase diagrams have been traditionally drawn from actual test data using the average gas velocity measured over a unit length of pipe as the abscissa. Because the gas is compressible, conditions are constantly changing throughout the entire length of the conveying line. In order to more correctly represent these conditions, a normalized phase diagram, where the dynamic pressure term is substituted for the velocity in the abscissa has been proposed (41).

A number of correlations for predicting saltation velocities in gas–solids flow appear in the literature, but apply only to a specific material and the specific conveying conditions under which they were measured (42). No general correlation has been found. All conveyor designers resort to testing to determine saltation velocities and conveying behavior. These data are then scaled to industrial size systems, using geometric and dynamic parameters that are based on experience.

Dilute Phase Conveying. Dilute conveying systems, sometimes called disperse conveying or stream conveying, operate as positive pressure systems at pressures up to 100 kPa (14.5 psig), or as negative pressure systems (vacuum conveying) at pressures up to –50 kPa (–500 mbar).

Calculation of the pressure drop in pneumatic conveyors is dependent on the use of experimentally derived correlations. Some success has been achieved in modeling the flow of well-dispersed, gas–solids mixtures, but these types of flow streams are not representative of most industrial pneumatic conveying situations. In an industrial plant, the most cost-effective design is one in which the solids are conveyed at the lowest velocity and highest solids loading consistent with maintaining a continuous, nonplugging stream. Under these conditions, the solids stream concentration is highest along the lower part of the horizontal pipes. This concentration changes throughout the length of the pipe, as solids are accelerated after the solids inlet, retarded then re-accelerated at bends in the pipe, and accelerated as the gas expands towards the terminal end of the conveying line. The gas-particle interactions in these nonhomogeneous flows have proven to be too complex to yield a reliable general correlation for calculating pressures, based on theory alone. Information on the application of two phase flow theory to pneumatic conveying design can be found in Reference 41.

Dense Phase Conveying. Dense phase conveying is often called high pressure conveying, or low velocity conveying. The term low velocity conveying is quite descriptive. The velocity pressure conditions that occur during dense phase conveying are shown in Figure 7. If the gas velocity is reduced below point D, the saltation velocity, the concentration of solids along the bottom of the pipe increases, reducing the gas flow area, causing a sudden increase in pressure and the onset of unstable flow conditions. Assuming the necessary pressure is available from the air source to continue to operate the system, the pressure follows the path D–E–F.

The mode of flow within the pipe is very sensitive to the characteristics of the material being conveyed. Materials that are permeable to air, eg, granular particles having a narrow particle size distribution, can move as naturally forming dunes or long plugs; materials that remain aerated for long periods of time can be transported as a moving bed; fine powders may form unstable moving beds that can become suddenly swept up and transformed into impermeable plugs that block the conveying pipe (43). For these reasons, the design and control of a dense phase conveying system requires close attention to the material handling characteristics. In many cases the individual systems are designed by the manufacturers to operate most efficiently when handling materials that exhibit specific flow characteristics. This is in contrast to dilute phase systems where a wider range of material characteristics can be accommodated by a single design arrangement.

Dense phase systems have been used to transport up to 50 t h. Although some very long systems up to 600 m have been reported, average systems range from about 150–300 m in length. Advantages over dilute phase systems are: low velocity resulting in more gentle handling of some materials; material can be conveyed over longer distances; the air requirements are less; and these conveyors are generally more energy efficient. Because less conveying air is used, smaller, less costly air solids separation and dust collection equipment are needed.

Dense phase systems are at cost disadvantage when multiple feed points are required, or when incremental compressed air for conveying is not available from existing sources. In the latter instance, the entire investment cost of an air compressor and associated auxiliaries must be allocated to the cost of the conveying system.

System Arrangements. Batch pressure tanks often called blow cases or pressure transporters, are commonly used to feed materials into dense phase conveying systems. These tanks have the capability to seal against the 100–800 kPa (1–8 bar) operating pressures required to operate these systems. Using a single tank feeder, the flow into the conveyor line is intermittent, as the control system cycles the tank operating sequence to fill, seal, pressurize, discharge, depressurize, and refill. Continuous flow can be achieved by using dual tank feeders operating on alternating discharge cycles.

Proprietary designs for rotary valve feeders (star valves) capable of continuous feeding of certain pelleted and granular materials into low velocity, dense phase systems, having system pressures up to 200 kPa (2 bars) have been developed.

Research and experimentation into the flow behavior of powders and granular materials have resulted in a number of commercial designs that can be customized to match the specific flow properties of the material, and can convey at controlled velocities and minimum plugging of the transport lines. Design improvements introduced include provision of means to introduce controlled amounts of air at appropriate points in the transport line, as well as the feed tank vessel, and introducing the air at timed (pulsed) intervals. These improvements have led to better control of conveying velocities, better control of dune and plug formation in the conveying lines with significant reduction in conveying line pluggages, reduced particle attrition, and reduced abrasive wear of the conveying line and fittings.

As in dilute phase conveying, very little published information on estimating pressure drop in dense phase conveyors is available. Industrial systems are designed in most part from scale-up of proprietary laboratory test data. Laboratory data gathered from many experiments using materials transported in 50 to 100 mm diameter dense phase conveying test rigs is available (44). The test methodology, and the techniques for scale-up of pressure transporter test rig data to industrial sized systems, should be of interest. Theory and applications of pneumatic conveying can be found in Reference 42 and information on conveying equipment and applications in References 44–48.

Economic Aspects

The bulk solids conveying industry is interwoven with all aspects of the chemical industry, from mining of raw materials to in-process handling and to final product delivery. The 1987 value of U.S. product shipments of bulk conveyors and parts was \$2.11 billion according to CEMA. Lists of conveyor manufacturers in the United States (49–51), and worldwide (52) are available.

BIBLIOGRAPHY

"Conveying" in *ECT* 1st ed., Vol. 4, pp. 347–379, by W. G. Hudson; in *ECT* 2nd ed., Vol. 6, pp. 92–121, by Richard R. Towers, Link-Belt Company; in *ECT* 3rd ed., Supplement Vol., pp. 282–309, by E. J. Anderson, Rexnord, Inc.

1. *Classification and Definitions of Bulk Materials*, CEMA 550, Conveyor Equipment Manufacturers Association, Rockville, Md., 1988.
2. A. W. Jenike, *Storage and Flow of Solids*, Bulletin 123, Engineering Experiment Station, University of Utah, 1964, rev. 1980.
3. *Pneumatic Handling of Powdered Materials*, Constable and Co., EEUA Handbook No. 15, London, 1963.
4. A. W. Roberts, *Bulk Solids Handling*, **11**, 17–35, (Mar. 1991).
5. J. W. Carson, *Bulk Solids Handling*, **11**, 37–44, (Mar. 1991).
6. L. Svarovsky, *Powder Testing Guide*, Elsevier, New York, 1987.
7. R. L. Brown and J. C. Richards, *Principles of Powder Mechanics*, Pergamon Press, Oxford, 1970; D. Mills, *Pneumatic Conveying Design Guide*, Butterworth, Boston, 1990.
8. J. Bridgewater in M. E. Fayed and L. Otten, eds., *Handbook of Powder Science and Technology*, Van Nostrand Reinhold, New York, 1984.
9. *Belt Conveyors for Bulk Materials*, 3rd ed., Conveyor Equipment Manufacturers Association, Rockville, Md., 1988.
10. *Continuous Mechanical Handling Equipment*, ISO 5048 International Standard, 1979.
11. *Bulk Conveyors for Bulk Materials*, DIN 22101 German Standard, 1982.
12. *Recommended Practice for Troughed Belt Conveyors*, The Mechanical Handling Engineers Association, London, 1986.
13. "Focus: Conveying Belt Technology," *Bulk Solids Handling* **2** (Nov. 1991).
14. *Belt Conveying Troughing and Return Idlers*, CEMA Standard 502-1980, Conveyor Equipment Manufacturers Association, Rockville, Md., 1980.
15. *Conveyor and Elevator Handbook*, 3rd ed., Rubber Manufacturers Association, Washington, D.C., 1988.
16. H. P. Lochmann, *Bulk Solids Handling* **4**, 841–849 (Oct. 1986).
17. J. W. Paelke, *Bulk Solids Handling* **10**, 273–280 (Aug. 1990).
18. E. Nissen, *Bulk Solids Handling* **11**, 899–901 (Nov. 1991).
19. J. C. Young in M. C. Hawk, ed., *Bulk Materials Handling*, Vol. 2, University of Pittsburgh, School of Engineering, 1973, pp. 238–361.
20. J. A. Dos Santos, *Bulk Solids Handling* **10**, 267–270 (Aug. 1990).
21. E. A. Bahke, *Bulk Solids Handling* **11**, 135–146 (Mar. 1991).
22. J. Patzke and K.-H. Kinzel, *Bulk Solids Handling* **10**, 305–308 (Aug. 1990).
23. R. Johanssen, *Bulk Solids Handling* **9**, 401–405 (Oct. 1989).
24. *Screw Conveyors*, CEMA Book No. 300, Conveyor Equipment Manufacturers Association, Rockville, Md., 1988.
25. *Screw Conveyor Dimensions*, CEMA Standard 300, Conveyor Equipment Manufacturers Association, Rockville, Md., 1981.
26. *Screw Conveyors-Trough Type for Industrial Use*, British Standard BS 4409, Part 1, British Standards Institution, London, 1969.
27. R. Rautenbach and W. Schumacher, *Bulk Solids Handling* **7**, 675–680 (Oct. 1987).
28. C.-Q. Mi, A. B. McLean, and P. C. Arnold, *Powder Handling & Processing* **3**, 227–239 (Sept. 1991).
29. *Ibid.*, 327–341 (Nov. 1991).

30.

F. V. Hetzel and R. K. Albright, *Belt Conveyors and Elevators*, John Wiley & Sons, Inc., New York, 1957.

31.

K. H. Koster, *Bulk Solids Handling* **5**, 449–460 (Apr. 1985).

32.

G. J. Beverdy, A. W. Roberts, and J. W. Hayes, *Bulk Solids Handling* **3**, 853–859 (Sept. 1983).

33.

H. Colijn, *Mechanical Conveying for Bulk Solids*, Elsevier, Amsterdam, 1985, pp. 236–328.

34.

H. A. Gaberson, *Trans. ASME Ser B* **94**(2), 50–63 (1972).

35.

Yigin Chu and Li Cuiying, *Bulk Solids Handling* **1**, 103–112 (Feb. 1987).

36.

C. R. Woodcock and J. S. Mason, *Bulk Solids Handling*, Chapman and Hall, New York, 1987, pp. 368–379.

37.

G. D. Dumbaugh, "A Comparative Review of Vibratory Drives for Bulk Solids Handling Systems," *Proceedings of the 10th Powder and Bulk Solids Conference*, Chicago, May 1985, pp. 425–470.

38.

R. E. Leitzel and W. M. Morrissey in Ref. 19, Vol. 1, 1971, pp. 307–325.

39.

C. R. Woodcock and J. S. Mason, *Bulk Solids Handling*, Chapman and Hall, New York, 1987, 456–476.

40.

F. A. Zenz and D. M. Othmer, *Fluidization and Particle Systems*, Reinhold, New York, 1960.

41.

R. D. Marcus, G. E. Klinzing, and F. Rizk, *Pneumatic Conveying of Solids*, Chapman and Hall, New York, 1991.

42.

M. V. Arastoopour, M. V. Modi, D. V. Punwani, and A. T. Talwalker, *A Review of Design Equations for Dilute Phase Gas-Solids Horizontal Conveying Systems for Coal and Related Material*, Powder and Bulk Solids Conference, Philadelphia, 1979.

43.

C. Y. Wen and H. P. Simon, *AIChE J.* **5**, 1969.

44.

D. Mills, *Pneumatic Conveying Design Guide*, Butterworths, New York, 1990.

45.

M. N. Kraus, *Pneumatic Conveying for Bulk Materials*, Prentice-Hall, Englewood Cliffs, N.J., 1991.

46.

H. A. Stoess, *Pneumatic Conveying*, Wiley-Interscience, New York, 1970.

47.

Plastics Pneumatic Conveying, G. Butters, ed. Applied Science Publishers, New Jersey, 1981.

48.

C. R. Woodcock and J. S. Mason, *Bulk Solids Handling*, Chapman and Hall, New York, 1987.

49.

Powder and Bulk Engineering 1992–1993 Resource and Buyers Guide (Updated annually), Minneapolis, MN.

50.

Powder/ Bulk Solids 1992 Annual Buyers Guide and Directory, Morris Plains, N.J.

51.

Chemical Processing Guide and Directory, Putnum Publishing Company, Chicago.

52.

1992 Year Book and Directory Powder & Bulk Solids Handling and Processing, Trans Tech Publications, Clausthal-Zellerfeld, Germany.

F. M. Thomson
CONSULTANT

COOLANTS.

See ANTIFREEZES AND DEICING FLUIDS; HEAT EXCHANGE TECHNOLOGY, HEAT TRANSFER MEDIA OTHER THAN WATER; REFRIGERATION.

COORDINATION COMPOUNDS

A coordination compound typically consists of a metal atom or ion (Lewis acid) surrounded by a number of electron-pair donors (Lewis bases) called ligands. Coordination compounds, also known as metal complexes, are pervasive throughout chemistry, biochemistry, and chemical technology. The metallic elements, which constitute 80% of the Periodic Table, exhibit predominantly coordination chemistry. Whereas the ligands may have charges equal in magnitude and opposite in sign to the charge of the metal ion, in which case a neutral coordination compound or inner complex results, often the metal plus ligands result in a charged entity or complex ion. In these situations, the coordination compound may include neutral ligands or counterions, either simple or complex. Generally coordination compounds have properties which are unique relative to both the ligands and the metal ion itself.

Coordination compounds are used as catalysts in nature, ie, metal enzymes, and in industry in situations in which the metal ion or ligand would not work alone. The ligands often modify the properties of the metals or metal ions. For example, the metal deactivators in gasoline modify the chemistry of dissolved copper to the extent that the copper does not promote gum formation (see GASOLINE AND OTHER MOTOR FUELS). Some ligands react with different metal ions such that selective analysis of metallic elements is possible through coordination followed by solvent extraction, spectroscopy, gravimetry, electrochemistry, etc. Furthermore, complexes of copper and zinc in water allow brass to be electroplated (see COPPER ALLOYS; ELECTROPLATING). Conversely, metals are sometimes used to modify the properties of ligands. For example, azo dyes (qv) are metallated to give more permanence and or alter color tones; the coordination of zinc to bactericides modifies the properties of the bactericides. The modification includes template and neighboring group effects, promotion of nucleophilic substitution, enhanced ligand acidity, and strain modification (1). Common geometries for coordination numbers from two through nine are shown in Figure 1, with IUPAC (2) and ACS (3) polyhedral symbols for the geometries provided in Table 1.

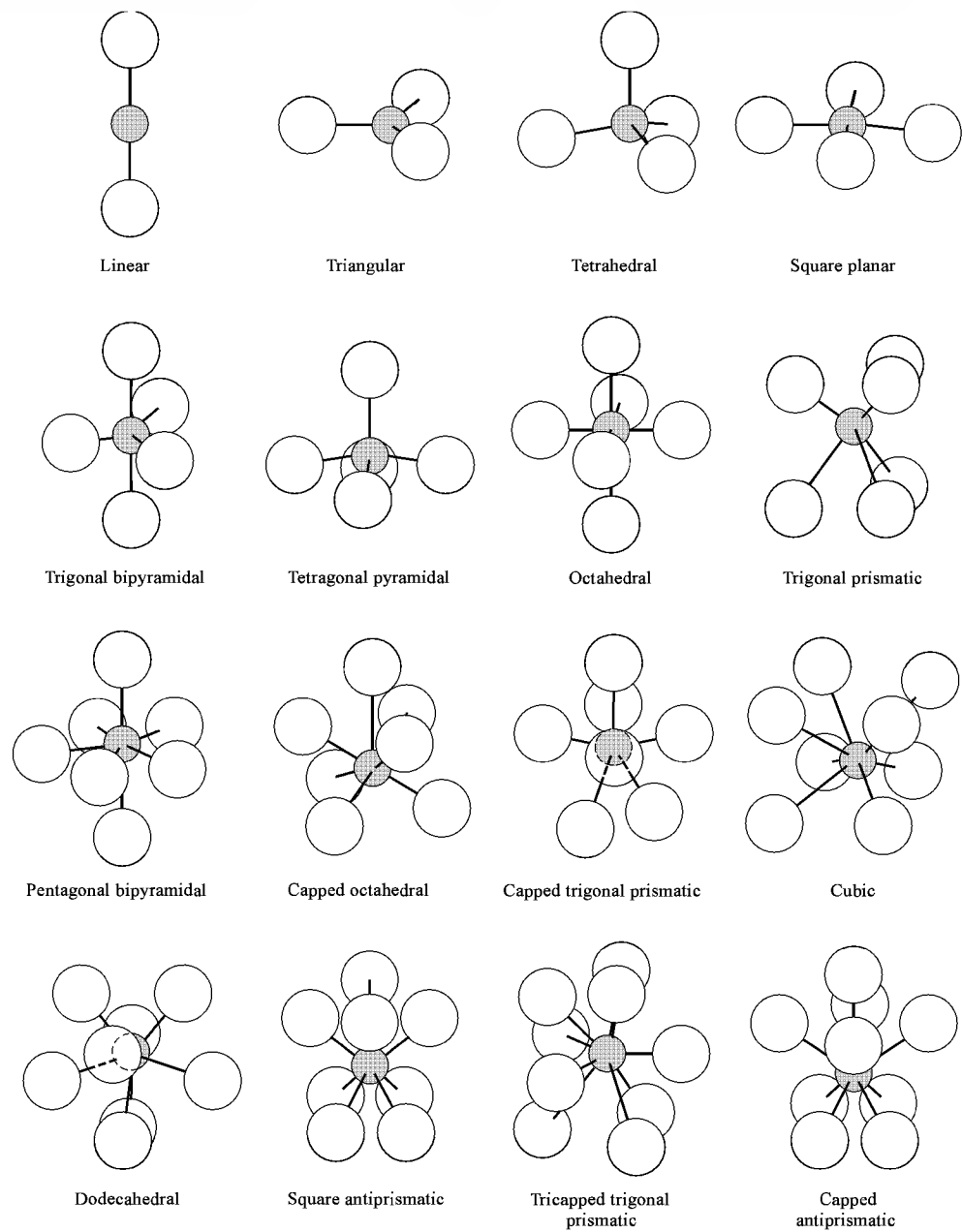


Fig. 1. Common geometries for metal coordination numbers from two through nine (see Table 1), where  represents the metal and , a ligand donor

atom. The principal axis orientation is vertical.

Table 1. Polyhedral Symbols for Common Coordination Geometries^a

Coordination geometry	Coordination number	Polyhedral symbols	
		IUPAC ^b	ACS ^c
linear	2	L-2	L-2
triangular plane	3	TP-3	TP-3
tetrahedron	4	T-4	T-4
square plane	4	SP-4	SP-4
trigonal bipyramid	5	TBPY-5	TB-5
square or tetragonal pyramid	5	SPY-5	SP-5
octahedron	6	OC-6	OC-6
trigonal prism	6	TPR-6	TP-6
pentagonal bipyramid	7	PBPY-7	PB-7
octahedron, face capped	7	OCF-7	OCF-7
trigonal prism, square face monocapped	7	TPRS-7	TPS-7
cubic	8	CU-8	CU-8
dodecahedral	8	DD-8	DD-8
square antiprism	8	SAPR-8	SA-8
trigonal prism, square face tricapped	9	TPRS-9	TPS-9

^a See Figure 1.
^b Ref. 2.
^c Ref. 3.

Nomenclature

Coordination compounds are named systematically beginning with the ligands. Ligands are listed alphabetically using prefixes that are Latin-based (di, tri, tetra, etc, for 2, 3, 4, etc) for identical ligands, or Greek-based (bis, tris, tetrakis, etc, for 2, 3, 4, etc) for identical multicomponent or complicated ligands that are placed in parentheses. The name of the metal follows, also using di, tri, etc prefixes if polynuclear, or ending in -ate if the overall species is anionic. Then in parentheses comes either the overall charge of the species, if any, using an Arabic number followed by a plus or minus sign (Ewens-Bassett nomenclature), or the oxidation number of the metal represented by a Roman numeral or Arabic zero (Stock nomenclature). Examples of simple coordination species, including names, are given in Table 2.

Table 2. Coordination Compounds Possessing Unidentate Ligands

Name ^a	CAS Registry Number	Molecular formula	Coordination number	Typical use
<i>Common coordination numbers</i>				
dicyanoargentate(I)	[15391-88-5]	[Ag(CN) ₂]	2	recovery of silver
tetraamminecopper(II)	[16828-95-8]	[Cu(NH ₃) ₄] ²⁺	4	dissolution of copper(II) in basic soln
tetracarbonylnickel(0)	[13463-39-3]	[Ni(CO) ₄]	4	separation of nickel from metals
hexafluorosilicate(IV)	[17084-08-1]	[SiF ₆] ²⁻	6	dissolution of silicon(IV) in hydro-lytic media
<i>Less common coordination numbers</i>				
tricyanocuprate(I)	[16593-63-8]	[Cu(CN) ₃] ²⁻	3	brass plating bath component
pentacarbonyl-manganate(-I)	[14971-26-7]	[Mn(CO) ₅]	5	composite polymer-metal particle synthesis
heptafluorozirconate(IV)	[27679-73-8]	[ZrF ₇] ³⁻	7	fluoride extractive metallurgy
octacyanotungstate(IV)	[18177-17-8]	[W(CN) ₈] ⁴⁻	8	light sensitizer for TiO ₂
nonaquaaneodymium(III)	[54375-24-5]	[Nd(H ₂ O) ₉] ³⁺	9	aqueous ion in Nd laser synthesis

^a Stock nomenclature

Bridging ligands are designated as μ or μⁿ, where *n* is the number of metal atoms bonded by the bridging ligand if *n* is greater than two. Such bridging ligands donate electron pairs to two or more metal ions simultaneously; eg, the W₂Cl₉³⁻ ion of K₃W₂Cl₉ has three bridging chloro ligands, and the potassium salt can be named as potassium tri-μ-chlorohexachloroditungstate(3-) [23403-17-0], K₃W₂Cl₉. Alternatively, the Ewens-Bassett (3-) can be replaced by (III) if the Stock nomenclature is used. The direct metal–metal bonded Re₂Cl₈²⁻ ion is named octachlorodirhenate(2-) (Re–Re) [19584-24-8] (or (III) may be used instead of (2-)). More details on nomenclature can be found in the literature (2).

Ligands having two or more donor atoms coordinated to the same acceptor metal atom or ion are called chelating ligands (see CHELATING AGENTS). The 1990 IUPAC *Nomenclature of Inorganic Chemistry* volume recommends that chelating ligands having 2, 3, 4, 5, 6, etc, donors bonded to the same metal atom be termed didentate, tridentate, tetradentate, pentadentate, hexadentate, etc, analogous to the prefixes used to designate identical ligands bonded to a metal center. This is in contrast to the bi-, ter-, quadra-, quinqi-, sexa-, etc, dentate of earlier IUPAC recommendations (4).

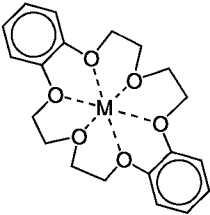
Other Types of Compounds

A coordination compound that has a chelating ligand in a ring around the metal ion is termed a macrocycle. Heme (1), the iron porphyrn derivative

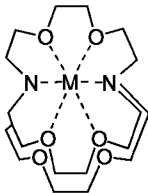
[14875-96-8] in human blood that carries oxygen, is an example. Other types of macrocyles such as dibenzo-[18]-crown-6 [14187-32-7] (2) known as crown ether and cryptate [23978-09-8] (3) complexes, as well as metal cluster coordination compounds, have been known for many years.



(1)

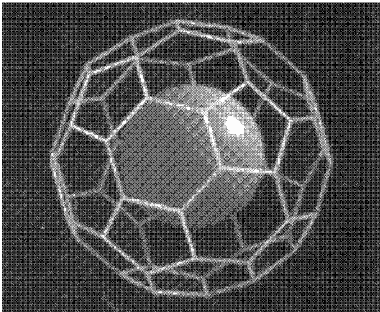


(2)

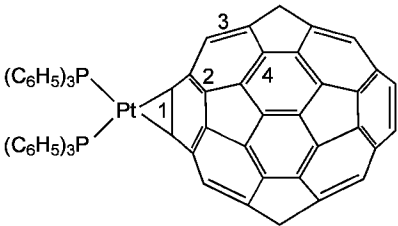


(3)

Newer are the metal fullerenes (5,6). A large number of metal encapsulated fullerenes, ie, three-dimensional C_x polytopal cage derivatives, have been observed, such as number of $M@C_{60}$ (4) species (see also CARBIDES). The @ indicates the metal is inside the C_{60} polytope, known as fullerene [99685-96-8]. There are also the smaller $M@C_{28}$ species for M = Ti, Zr, Hf, and U, that are thought to be stabilized by these tetravalent metals. External fullerene complexes, eg, $[(C_6H_5)_3P]_2Pt(\eta^2-C_{60})$ [135863-99-9] (5) and $\{[(C_2H_5)_3P]_2Pt\}(\eta^2-C_{60})$ [137233-82-0] are also known where the η^2 indicates that two of the carbons in the C_{60} are bonded to the platinum. Additionally, alkali metal doped fullerenes having several metal atoms per fullerene show interesting properties. For example, K_3C_{60} [137232-17-8] and Rb_3C_{60} [137926-73-9] become superconducting at 18 K and 28 K, respectively. Another class of three-dimensional structures are the icosahedral metallocarbohedranes such as M_8C_{12} , where M = Ti, Zr, or Hf.



(4)



(5)

Whereas dihydrogen was formerly thought to be bonded to metals only as the dissociated hydrido (H^-) species, a number of η^2-H_2 complexes are known (7). The η^2-H_2 designation implies the $H-H$ bond is still intact and both hydrogens are coordinated to the metal in a manner analogous to the η^5 -cyclopentadienyl organometallic ligands exemplified by ferrocene [102-54-5]. Discovery of these dihydrogen species altered the thinking of chemists not only with regard to hydrogenation processes but even regarding the very nature of bonding between metals and ligands (7). Metals bonded to hydrocarbons through carbon atoms are normally termed organometallics. Only the mixed ligand species where C-bonded ligands are but one of the ligand types are discussed herein.

Coordination Theories and Bonding

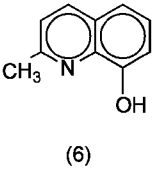
The quest for a comprehensible theory of coordination chemistry has given rise to the use of valence-bond, crystal-field, ligand-field, and molecular-orbital

Al ³⁺ , Ga ³⁺ , In ³⁺ Si ⁴⁺ , Sn ²⁺ , 4+, As ³⁺ metals in oxidation state 4+ or higher ^b	Ni ²⁺ , Cu ²⁺ Zn ²⁺ Pb ²⁺	Pd ²⁺ , Pt ²⁺ , Cu ⁺ , Ag ⁺ , Au ^{+,3+} Cd ²⁺ , Hg ₂ ²⁺ , Hg ²⁺ , Tl ^{+,3+}
metal in oxidation state 0 or lower		
<i>Ligands</i>		
F, O donors		H, CO, CO, R alkide ions, olefins aromatics
	N donors Cl, Br	P, As, Sb, Bi donors S, Se, Te donors, I
<i>Solvents</i>		
HF, H ₂ O, ROH	NH ₃	(CH ₃) ₂ CO, (CH ₃) ₂ SO, (CH ₂) ₄ SO ₂ , nitroparaffins, DMF

^a DMG = dimethylglyoxime; DMF = dimethyl formamide.
^b Exceptions exist if a large number of soft bases are coordinated to the metal.

Chelating or multidentate ligands are usually more stable than the analogous unidentate ligands, and this gives rise to the chelate effect. This effect is particularly evident for the formation of five-membered rings, anionic ligands, and aromatic donors. Larger rings, unless rigid, lose appreciable entropy on coordination; anionic ligands displace large quantities of oriented solvent molecules; and aromatic ring donors are normally rigid. Detailed quantitative considerations between similar systems is often difficult, however, because the Δ*H* and Δ*S* changes are normally a very small fraction of the enthalpies involved in the coordination bonds.

Steric Selectivity. In addition to the normal regularities that can be rationalized by electronic considerations, steric factors are important in coordination chemistry. To illustrate, 8-hydroxyquinoline, or 8-quinolinol (Hq) [148-24-3], at 100°C precipitates both Mg²⁺ and Al³⁺ from aqueous solution as hydrated Mg(q)₂ (formulated as Mg(q)₂(H₂O)₂ [56531-18-1]) and as Al(q)₃ [2085-33-8], respectively. 2-Methyl-8-hydroxyquinoline [826-81-3] (6),



however, precipitates only Mg²⁺ from aqueous solution. The 2-methyl group prevents three of the methyl-substituted ligands from coordinating to Al³⁺ in water. By varying ring sizes and donor atoms of the macrocyclic crown ethers and the fused ring cryptates (or cryptards), different metal ions can be selectively accommodated. The crown ethers, such as 18-crown-6 o eicosahydrodibenzo[*b,k*][1,4,7,10,13,16]hexaoxacyclooctadecin [14187-32-7] (2) are highly selective for K⁺ relative to Na⁺. However, the cryptate 4,7,13,16,21,24-hexaoxa-1,10-diazabicyclo[8.8.8]hexacoxane [23978-09-8] (3) is selective for the sodium ion (see ATALYSIS, PHASE TRANSFER).

Another example of steric selectivity involves the homopoly and heteropoly ions of molybdenum, tungsten, etc. Each molybdenum(VI) and tungsten(VI) ion is octahedrally coordinated to six oxygen (oxo) ligands. Chromium(VI) is too small and forms only the well-known chromate-type species having four oxo ligands. The ability of other cations to participate in stable heteropoly ion formation is also size related.

Coordination stereochemistry (including various forms of isomerization) is an area of significant research interest. This aspect of coordination is important for stereospecific catalytical applications.

Reactions

Substitution. Coordination species are often categorized in terms of the rate at which they undergo substitution reactions. Complexes that react with other ligands to give equilibrium conditions almost as fast as the reagents can be mixed by conventional techniques are termed labile. Included are most of the complexes of the alkali metals, the alkaline earths, the aluminum family, the lanthanides, the actinides, and some of the transition-metal complexes. On the other hand, numerous transition-metal complexes that kinetically resist substitution reactions are termed inert. These terms refer to substitution reactivity and not to thermodynamic properties. As an illustration, the reaction rates, *k*, and activation parameters, Δ*H*^{*}, Δ*S*^{*}, and Δ*V*^{*}, of solvent-exchange for some of the aquated cations of the first-row transition elements are listed in Table 4. Although vanadium(II) and nickel(II) react more slowly than the other ions of oxidation state II listed, they are labile by the operational definition. On the other hand, chromium(III) and cobalt(III) complexes are inert. Iron(III), however, is labile even though iron lies between chromium and cobalt in the Periodic Table.

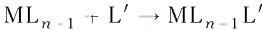
Table 4. Water Exchange Rates and Activation Parameters of Hexaaqua Complexes at 25°C^{a, b}

Ion	<i>k</i> ₁ , s ⁻¹	Δ <i>H</i> [*] ,	Δ <i>S</i> [*] , J (K·mol) ^c	Δ <i>V</i> [*] ,	Electron configuration ^d
		kJ mol ^c		cm ³ mol	
V ²⁺	8.9 × 10 ¹	62	0.4	4.1	<i>d</i> ³
Cr ²⁺	7 × 10 ⁹				<i>d</i> ⁴
Mn ²⁺	2.1 × 10 ⁷	33	6	5.4	<i>d</i> ⁵
Fe ²⁺	4.4 × 10 ⁶	41	21	3.8	<i>d</i> ⁶
Co ²⁺	3.2 × 10 ⁶	47	37	6.1	<i>d</i> ⁷
Ni ²⁺	3.2 × 10 ⁴	57	32	7.2	<i>d</i> ⁸
Cu ²⁺	8 × 10 ⁹	20			<i>d</i> ⁹
Ti ³⁺	1.8 × 10 ⁶	43	1	12.1	<i>d</i> ¹
V ³⁺	5.0 × 10 ²	49	28	8.9	<i>d</i> ²
Cr ³⁺	2.4 × 10 ⁻⁶	109	12	9.6	<i>d</i> ³
Fe ³⁺	1.6 × 10 ²	64	12	-5.4	<i>d</i> ⁵
CrOH ²⁺	1.8 × 10 ⁻⁴	110	55	2.7	<i>d</i> ³
FeOH ²⁺	1.2 × 10 ⁶	42	5	7.0	<i>d</i> ⁵

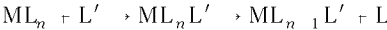
^a Ref. 1, unless noted otherwise.
^b For the reaction M(H₂O)₆ⁿ⁺ + H₂¹⁸O → M(H₂O)₅(H₂¹⁸O)ⁿ⁺ + H₂O.
^c To convert from J to cal, divide by 4.184.
^d The *d*⁴ (Cr²⁺) and *d*⁹ (Cu²⁺) complexes have distorted octahedra.

In general, octahedral complexes of transition-metal ions possessing 0, 1, or 2 electrons beyond the electronic configuration of the preceding noble gas, ie, d^0 , d^1 , d^2 configurations, are labile. The d^8 systems are usually inert; the relative lability of vanadium(II) may be charge and/or redox related. However, high spin d^4 , d^5 , and d^6 species, which possess 4, 5, and 4 unpaired electrons, respectively, are labile, as are d^7 through d^{10} octahedral complexes. In addition to the inert d^8 systems, low spin d^4 , d^5 , and d^6 complexes are inert to rapid substitution. The d^8 species are the least labile of the configurations classed as labile.

Spin-paired octahedral d^8 ions and spin-free octahedral d^8 ions appear to react by largely dissociative (D or I_d) reactions, for example,



although d^8 complexes which are just barely spin-paired, such as tri-1,10-phenanthrolineiron(II) [14708-99-7] and ethylenediaminetetraacetatocobalt(III) [15136-66-0], appear to undergo associative (A or I_a) reactions, with good nucleophiles.



Discernible associative character is operative for divalent $3d$ ions through manganese and the trivalent ions through iron, as is evident from the volumes of activation in Table 4. However, deprotonation of a water molecule enhances the reaction rates by utilizing a conjugate base π -donation dissociative pathway. As can be seen from Table 4, there is a change in sign of the volume of activation $\Delta V^\ddagger$. Four-coordinate square-planar molecules also show associative behavior in their reactions.

For many species the effective atomic number (EAN) or 18-electron rule is helpful. Low spin transition-metal complexes having the EAN of the next noble gas (Table 5), which have 18 valence electrons, are usually inert, and normally react by dissociation. Each normal donor is considered to contribute two electrons; the remainder are metal valence electrons. Sixteen-electron complexes are often inert, if these are low spin and square-planar, but can undergo associative substitution and oxidative-addition reactions.

Table 5. Complexes Having Effective Atomic Numbers (EAN) of a Noble Gas^a

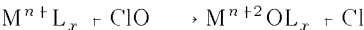
Name	CAS Registry Number	Molecular formula	Coordination number	EAN
nickel tetracarbonyl	[13463-39-3]	Ni(CO) ₄	4	36
iron pentacarbonyl	[13463-40-6]	Fe(CO) ₅	5	36
chromium hexacarbonyl	[13007-92-6]	Cr(CO) ₆	6	36
hexaammineplatinum(IV)	[16893-12-2]	Pt(NH ₃) ₆ ⁴⁺	6	86
tricarbonyldichlorobis-(triphenylphosphine)-molybdenum(II)	[17250-39-4]	Mo(CO) ₃ Cl ₂ [P(C ₆ H ₅) ₃] ₂	7	54
tetrakis(8-quinolino-lato) tungsten(IV)	[17499-74-0]	W(C ₉ H ₇ NO) ₄	8	86
nonahydridorhenate(VII)	[44863-47-0]	[ReH ₉] ²⁻	9	86

^a 18-Electron species.

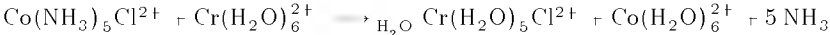
Oxidation–Reduction. Redox or oxidation–reduction reactions are often governed by the hard–soft base rule. For example, a metal in a low oxidation state (relatively soft) can be oxidized more easily if surrounded by hard ligands or a hard solvent. Metals tend toward hard-acid behavior on oxidation. Redox rates are often limited by substitution rates of the reactant so that direct electron transfer can occur (16). If substitution is very slow, an outer sphere or tunneling reaction may occur. One-electron transfers are normally favored over multielectron processes, especially when three or more species must aggregate prior to reaction. However, oxidative addition



and oxygen atom transfer

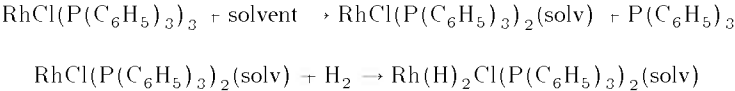


where M is a metal ion of charge $n+$ and L is a ligand, are considered complementary two-electron transfers that are second only to one-electron processes in general. The distinction between atom and electron transfer is academic when a bridging atom moves from one species to another during a redox reaction:



In effect, the atom is transferred.

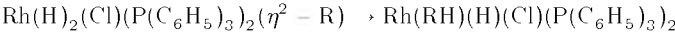
Coordination species which coordinate an organometallic ligand in catalytic processes often cycle between 16- and 18-electron systems, although ligand substitution also participates. For example, the first step in the well-studied olefin hydrogenation by chlorotris(triphenylphosphine)rhodium(I) [14694-95-2] is now thought to be solvation followed by oxidative addition of hydrogen because the phosphine is somewhat labile and the solvated species adds hydrogen about 10⁷ times as fast as the original complex and goes from a 16-electron rhodium(I) species to an 18-electron rhodium(III) species (17):



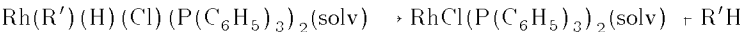
where solv is solvent. An olefin, R, such as cyclohexene, substitutes for the solvent on the dihydride.



This equilibrium is followed by a rate-determining tautomeric shift (18 → 16 electrons) to give a coordinated alkyl group:



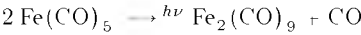
where RH = alkyl. These last two steps are often called an olefin insertion. Addition of solvent gives the 18-electron complex, Rh(R')(H)(Cl)(P(C₆H₅)₃)₂(solv), and reductive elimination produces the original 16-electron solvated catalyst:



This catalytic cycle is related to some stereoselective industrial catalysis.

Photochemistry. Substitution rates of many complexes are enhanced by irradiation of the low energy $d-d$ transitions, such as $t_{2g} \rightarrow e_g$ in octahedral coordination compounds (see Fig. 2). Quantum yields, Φ, defined as the ratios of moles of product formed (or reactant depleted) to the moles

of photons absorbed, vary from very good, eg, chromium(III) ca 0.5, to poor, eg, cobalt(III) <0.01, for ligand substitution (18). The substituted ligand is normally the stronger ligand of the two on the axis with the weakest net pair of ligands as determined by spectrochemical relationships, ie, $\text{CN}^- > \text{NO}_2^- > \text{NH}_3 > \text{H}_2\text{O}, \text{F}^- > \text{Cl}^-$. Exceptions do occur. Photochemical ligand dissociation is useful in the synthesis of multinuclear metal complexes such as diiron nonacarbonyl [15321-51-4] from iron pentacarbonyl [13463-40-6]

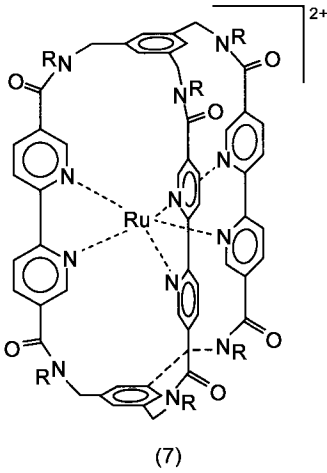


Or conversely, active radicals can be obtained by irradiating certain metal—metal bonded species:



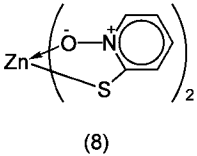
Irradiation of coordination compounds in the charge-transfer spectral region can often enhance redox reactions. The quantum yields are variable.

The use of photochemical redox for practical energy-transfer is being actively pursued (see HYDROGEN; HYDROGEN ENERGY). For example, sufficient energy is available in visible photons to split water into hydrogen and oxygen (19). Whereas derivatives of $\text{Ru}(\text{bpy})_3^{2+}$ [15158-62-0], where bpy 2,2'-bipyridine, have been at the forefront of this area, a small amount of decomposition during the photochemical cycles had precluded commercialization of such systems. However, the encapsulated ruthenium(II) [116970-07-1] found in structure (7), where R is benzyl, is claimed to be 10^4 times more stable than $\text{Ru}(\text{bpy})_3^{2+}$ (20).



Applications

Bactericides and Fungicides. Among marketed bactericides is the sodium salt of 2-pyridinethiol-1-oxide [3811-73-2], $\text{C}_5\text{H}_5\text{NOSNa}$. However, because this material is a strong skin irritant, the milder zinc chelate [13463-41-7] (8) is used in shampoos (see ANTIBACTERIAL AGENTS, SYNTHETIC; HAIR PREPARATIONS) and also has been used in wood protection. The plant fungicides zineb [2122-67-7], $\text{Zn}(\text{S}_2\text{CNHCH}_2\text{CH}_2\text{HCS}_2)_2$, and maneb [12427-38-2], $\text{Mn}(\text{S}_2\text{CNHCH}_2\text{CH}_2\text{NHCS}_2)_2$, are polymeric chelates (see FUNGICIDES, AGRICULTURAL). A number of other coordination compounds, especially copper derivatives, are used as bactericides, fungicides, and disinfectants (21) (see DISINFECTANTS AND ANTISEPTICS). Some chelating agents used in the past, eg, ligands derived from 8-quinolinol, are no longer in use because of concern over their ability to facilitate unwanted metal ion transport.



Catalysis. Hydrogenation reactions can be catalyzed by a wide variety of coordination compounds (22) (see CATALYSIS). For example, dicobalt octacarbonyl [10210-68-1], $\text{Co}_2(\text{CO})_8$, has been suggested to be a suitable catalyst for the hydrogenation of olefins to alkanes, aldehydes to alcohols, acid anhydrides to acids and aldehydes, as well as for the selective hydrogenation of polyenes to monoenes, and for the hydrogenation and isomerization of unsaturated fats (see also OXO PROCESS). The list for chlorotris(triphenylphosphine)rhodium(I) [14694-95-2], Wilkinson's catalyst is even longer (23). On the other hand, a number of chromium tricarbonyl arene derivatives such as benzenetricarbonylchromium [12082-08-5], $\text{C}_6\text{H}_6\text{Cr}(\text{CO})_3$, appear to be quite selective for the hydrogenation of 1,3- and 1,4-dienes to form *cis*-monoenes by 1,4-addition. Other homogeneous hydrogenation catalysts include coordination compounds of Fe, Ru, Ir, Ni, Pd, Pt, Os, and Cu. Probably the most important homogeneous hydrogenation application has been the Monsanto asymmetric (or enantioselective) synthesis of L-dopa [59-92-7], $\text{C}_9\text{H}_{11}\text{NO}_4$, a chiral amino acid used to treat Parkinson's disease, using a rhodium(I) catalyst containing a chiral phosphine (24).

The palladium chloride process for oxidizing olefins to aldehydes in aqueous solution (Wacker process) apparently involves an intermediate anionic complex such as dichloro(ethylene)hydroxopalladate(II) or else a neutral aqua complex $\text{PdCl}_2(\text{CH}_2=\text{CH}_2)(\text{H}_2\text{O})$. The coordinated PdCl_2 is reduced to Pd during the olefin oxidation and is reoxidized by the cupric-cuprous chloride couple, which in turn is reoxidized by oxygen, and the net reaction for any olefin ($\text{RCH}=\text{CH}_2$) is then



Whereas this reaction was used to oxidize ethylene (qv) to acetaldehyde (qv), which in turn was oxidized to acetic acid, the direct carbonylation of methanol (qv) to acetic acid has largely replaced the Wacker process industrially (see ACETIC ACID AND DERIVATIVES). A large number of other oxidation reactions of hydrocarbons by oxygen involve coordination compounds as detailed elsewhere (25).

The oxo reaction or hydroformylation converts an olefin to an aldehyde containing one or more carbon atoms, using tetracarbonylhydrocobalt [16842-03-8], $\text{CoH}(\text{CO})_4$. Phosphine substitution lowers the pressure required for the hydroformylation. Rhodium coordination species are also used for this purpose and give a higher normal to branched aldehyde ratio in the hydroformylation of alkenes, and both rhodium and platinum species appear useful for asymmetric hydroformylation (26). The carbonylation of methanol by CO to acetic acid is catalyzed by coordination compounds of all three Group 9 metals (Co, Rh, and Ir); however, the Monsanto process using rhodium complexes allows the use of lower CO pressures and is used worldwide. Adiponitrile is prepared industrially by the hydrocyanation (addition to two moles of HCN) to butadiene (qv) using a $\text{Ni}[\text{P}(\text{OC}_2\text{H}_5)_3]_4$ catalyst.

Metal coordination compounds may also provide alternatives to the heterogeneous catalysts used for the water gas shift reaction. In fact, Ru, Rh, Ir, and Pt coordination compounds have all shown some promise (27).

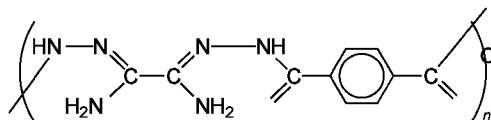
The stereospecific polymerization of alkenes is catalyzed by coordination compounds such as Ziegler-Natta catalysts, which are heterogeneous TiCl_3 -Al alkyl complexes. Cobalt carbonyl is a catalyst for the polymerization of monoepoxides; several rhodium and iridium coordination compounds

catalyze certain polymerizations; and nickel complexes can initiate living polymers. Cyclooligomerization is promoted by coordination compounds of nickel, as well. The products depend on the sites available for coordination; eg, complexes having four octahedral sites available for coordination by acetylene produce cyclooctatetraene, three facial sites yield benzene, etc. Other transition-metal species catalyze similar cyclizations to benzene.

Olefin isomerization can be catalyzed by a number of catalysts such as molybdenum hexacarbonyl [13939-06-5], Mo(CO)_6 . This compound has also been found to catalyze the photopolymerization of vinyl monomers, the cyclization of olefins, the epoxidation of alkenes and peroxo species, the conversion of isocyanates to carbodiimides, etc. Rhodium carbonylhydrottris(triphenylphosphine) [17185-29-4], $\text{RhH(CO)(P(C}_6\text{H}_5)_3)_3$, is a multifunctional catalyst which accelerates the isomerization and hydroformylation of alkenes.

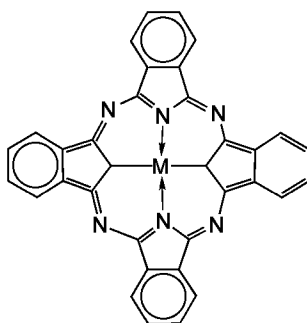
These applications are mostly examples of homogeneous catalysis. Coordination catalysts that are attached to polymers via phosphine, siloxy, or other side chains have also shown promise. The catalytic specificity is often modified by such immobilization. Metal enzymes are, from this point of view, anchored coordination catalysts immobilized by the protein chains. Even multistep syntheses are possible using alternating catalysts along polymer chains. Other polynuclear coordination species, such as the homopoly and heteropoly ions, also have applications in reaction catalysis.

Coordination Polymers. In addition to catalysts, polymeric coordination compounds have applications in high temperature coatings (qv) and lubricants utilizing poly(metal phosphinates) (see LUBRICATION AND LUBRICANTS), and also in the chelated fiber called Enkatherm, which is the zinc chelate of poly(terephthaloyl oxalic-bis-amidrazone), PTO (9). Enkatherm has been said to resist temperatures of up to 1500°C, and to be superior to any other flame-resistant fiber (28) (see FLAME RETARDANTS FOR TEXTILES). Sensitivity to acid solutions, poor long-term stability, and photosensitivity have limited the use of such materials. On the other hand, there are a large number of potential uses for metal coordination polymers (29), including use as high energy lithographic resists (see LITHOGRAPHIC RESISTS) (30).

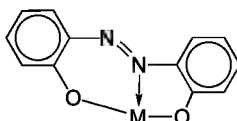


(9)

Dyes and Pigments. Several thousand metric tons of metallated or metal coordinated phthalocyanine dyes (10) are sold annually in the United States. The partially oxidized metallated phthalocyanine dyes are good conductors and are called molecular metals (see SEMICONDUCTORS; PHthalocyanine compounds; COLORANTS FOR PLASTICS). Azo dyes (qv) are also often metallated. The basic unit for a 2,2'-azobisphenol dye is shown as structure (11). Sulfonic acid groups are used to provide solubility, and a wide variety of other substituents influence color and stability. Such complexes have also found applications as analytical indicators, pigments (qv), and paint additives.



(10)

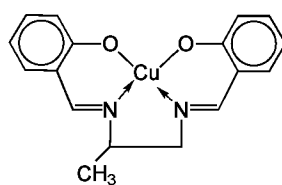


(11)

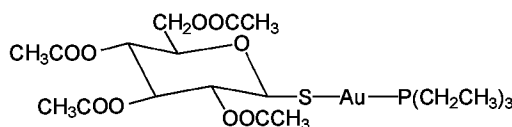
Photography. Coordination chemistry is important in classical black and white silver photographic systems, color photography (qv), and in other special photographic areas (see PHOTOGRAPHY) (31). In order to have silver halide grains of appropriate size, silver halides are ripened by solutions containing a silver complexing agent. For example, $[\text{AgX}_2]^-$, where X is a halide, or $[\text{Ag}(\text{NH}_3)_2]^+$ [16972-61-5] can be used to partially dissolve and, through dilution, reprecipitate the silver halides in the gelatin emulsion. In fact, a light-insensitive emulsion consisting of $(\text{NH}_4)_2[\text{AgBr}_3]$ becomes light-sensitive AgBr when moistened with water or alcohol. Silver halides have increased light sensitivity when sensitized using 10^{-5} moles of a gold(I) or gold(III) coordination compound, such as ammonium dithiocyanatoaurate(I) [15066-31-6], $\text{NH}_4[\text{Au}(\text{SCN})_2]$, or sodium tetrachloroaurate(III) [15189-51-2], $\text{Na}[\text{AuCl}_4]$, per mole of silver halide. Coordination compounds are used as emulsion stabilizers, developers, and are formed with the well-known thiosulfate fixers. Silver halide diffusion transfer processes and silver image stabilization also make use of coordination phenomena. A number of copper and chromium azo dyes have found use in diffusion transfer systems developed by Polaroid (see COLOR PHOTOGRAPHY, INSTANT). Coordination compounds are also important in a number of commercial photothermography and electrophotography (qv) applications as well as in the classic iron cyano blueprint images, a number of chromium systems, etc (32).

Electroplating. Aluminum can be electroplated by the electrolytic reduction of cryolite, which is trisodium aluminum hexafluoride [13775-53-6], Na_3AlF_6 , containing alumina. Brass (see COPPER ALLOYS) can be electroplated from aqueous cyanide solutions which contain cyano complexes of zinc(II) and copper(I). The soft CN^- stabilizes the copper as copper(I) and the two cyano complexes have comparable potentials. Without CN^- the potentials of aqueous zinc(II) and copper(I), as well as those of zinc(II) and copper(II), are over one volt apart; thus only the copper plates out. Careful control of concentration and pH also enables brass to be deposited from solutions of citrate and tartrate. The noble metals are often plated from solutions in which coordination compounds help provide fine, even deposits (see ELECTROPLATING).

Petroleum Additives. Antiknock fuel additives include lanthanide β -diketones, several mixed-ligand manganese carbonyl complexes, and tetraethyllead. Coordination compounds have been suggested as fuel oil additives. Metal deactivators for gasoline include Schiff-base ligands that minimize oxidation state changes in the traces of copper dissolving from fuel lines. For example, the planar copper(II) chelate [14522-52-2] (12) resists reduction to copper(I) in gasoline (see GASOLINE AND OTHER MOTOR FUELS).



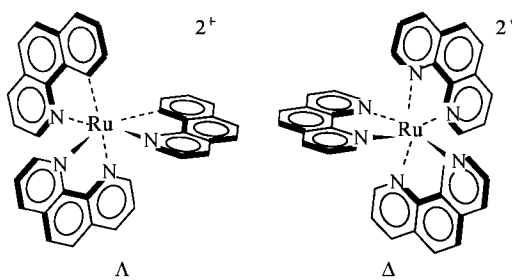
(12)



(13)

Therapeutic Chelates. Calcium ethylenediaminetetraacetate [38620-52-9] is used to treat lead poisoning; the free ligand causes calcium loss. (see LEAD COMPOUNDS, TOXICOLOGY). Platinum coordination compounds, such as cisplatin, *cis*-dichlorodiammineplatinum(II) [15663-27-1], *cis*-Pt(NH₃)₂Cl₂, are cancer therapeutics (33) (see CHEMOTHERAPEUTICS, ANTICANCER). A number of other neutral coordination compounds of platinum(II) and (IV), and of other metals, are also under investigation. The improved cancer cure rates using combined chemotherapy and radiotherapy is leading to the development of Auger-emitting therapeutic drugs. Gold complexes, such as (13) [34031-32-8], have been used to treat rheumatoid arthritis since 1935, and copper complexes may also be of some benefit as antiinflammatory drugs. There are many other examples available in the literature (34).

Technetium-99m coordination compounds are used very widely as noninvasive imaging tools (35) (see IMAGING TECHNOLOGY; RADIOACTIVE TRACERS). Different coordination species concentrate in different organs. Several of the [Tc^VO(chelate)₂][±] types have been used. In fact, the large majority of nuclear medicine scans in the United States are of technetium-99m complexes. Moreover, chiral transition-metal complexes have been used to probe nucleic acid structure (see NUCLEIC ACIDS). For example, the two chiral isomers of tris(1,10-phenanthroline)ruthenium(II) [24162-09-2] (14) interact differently with DNA. These compounds are enantioselective and provide an addition tool for DNA structural interpretation (36).



Miscellaneous. Numerous other applications of coordination compounds exist and may be found throughout the *Encyclopedia*. Tetrahedral cobalt(II) units are incorporated in cobalt blue glass (see COBALT AND COBALT ALLOYS; CERAMICS; GLASS). Tetrahedral to octahedral coordination and the concurrent blue to pink indicator color change is obtained from moisture on cobalt(II) chloride (see DESICCANTS; DRYING). Nickel cyano clathrates are used for organic isomer separation, and the closely related zeolites as catalysts (37) and as porous molecular sieves (qv) (see INCLUSION COMPOUNDS). A number of dioxygen complexes are being considered for artificial blood applications (see BLOOD, ARTIFICIAL); tetrakis(β-diketonato) anionic chelates of europium are employed in laser applications (see LASERS); the neutral tris(β-diketonato) lanthanide chelates are used as nmr shift reagents (see MAGNETIC SPIN RESONANCE); and coordination compound membranes are under consideration for ion-selective electrodes (see ELECTROANALYTICAL TECHNIQUES). The use of coordination compounds in extractive metallurgy includes complexes in leaching, solvent-extraction, ion-exchange (qv), precipitation, and electrochemical processes (see METALLURGY, EXTRACTIVE) (38). Similarly, nuclear fuel cell preparation and reprocessing use coordination compounds (39) (see NUCLEAR REACTORS, NUCLEAR FUEL).

BIBLIOGRAPHY

"Coordination Compounds" in *ECT* 1st ed., Vol. 4, pp. 379–391, by R. D. Johnson, The University of Pittsburgh; in *ECT* 2nd ed., Vol. 6, pp. 122–131, by N. Christian Nielsen, Central Methodist College; in *ECT* 3rd ed., Vol. 6, pp. 784–798 by R. D. Archer, University of Massachusetts.

1. R. G. Wilkins, *Kinetics and Mechanism of Reactions of Transition Metal Complexes*, VCH, Weinheim, Germany, 1991, Chapt. 6.
2. G. J. Leigh, ed., *Nomenclature of Inorganic Chemistry, Recommendations 1990*, International Union of Pure and Applied Chemistry, Blackwell Scientific Publications, Oxford, UK, 1990.
3. B. P. Block, W. H. Powell, and W. C. Fernelius, *Inorganic Chemical Nomenclature, Principles and Practice*, American Chemical Society, Washington, D.C., 1990, pp. 144–;147.
4. *IUPAC Nomenclature of Inorganic Chemistry*, 2nd ed., Butterworths, London, 1971.
5. R. M. Baum, *Chem. Eng. News* **70**(22), 25–27, 30–33 (June 1, 1992). Reviews advances in fullerene chemistry including metal coordination, but without a true bibliography.
6. *Acc. Chem. Res.* **25**, 98–175 (Mar. 1992). A special issue on fullerenes, small three-dimensional carbon cages such as C₆₀, although many other C_n are now known as well, including their metal derivatives.
7. G. J. Kubas, *Acc. Chem. Res.* **21**, 120–128 (1988).
8. F. A. Cotton, *Chemical Applications of Group Theory*, 3rd ed., Wiley-Interscience, New York, 1990; S. F. A. Kettle, *Symmetry and Structure*, John Wiley & Sons, Chichester, UK, 1985.
9. G. Wilkinson, R. D. Gillard, and J. A. McCleverty, eds., *Comprehensive Coordination Chemistry*, Vol. 1, Pergamon Press, Oxford, UK, 1987.
10. R. DeKock and H. B. Gray, *Chemical Structure and Bonding*, Benjamin-Cummings, Menlo Park, Calif., 1980; F. Basolo and R. Johnson, *Coordination Chemistry*, Science Reviews, Northwood, UK, 1987. For more detailed treatments, see A. F. Williams, *A Theoretical Approach to Inorganic Chemistry*, Springer-Verlag, Berlin, 1979; J. S. Griffith, *The Theory of Transition Metal Ions*, Cambridge University Press, UK, 1961; M. Gerloch, *Magnetism and Ligand-Field Analysis*, Cambridge University Press, UK, 1983.
11. F. A. Cotton and G. Wilkinson, *Advanced Inorganic Chemistry*, 5th ed., John Wiley & Sons, Inc., New York, 1988, Chapt. 23.
12. Vol. 1, "Amino Acids," 1974; Vol. 2, "Amines," 1975; Vol. 3, "Other Organic Ligands," 1977; Vol. 4, "Inorganic Complexes," 1976; and Vol. 5, "Supplement," 1982, in A. E. Martell and R. M. Smith, eds., *Critical Stability Constants*, Plenum Press, New York.
13. S. Ahlrand, J. Chatt, and N. R. Davies, *Quart. Rev.* **11**, 265 (1958); S. Ahlrand, *Struct. Bonding (Berlin)* **5**, 118 (1968).

14. R. G. Parr and R. G. Pearson, *J. Am. Chem. Soc.* **105**, 7512 (1983); R. G. Pearson, *J. Chem. Educ.* **64**, 561 (1987).
15. H. Irving and R. J. P. Williams, *Nature (London)* **162**, 746 (1948); H. Irving and R. J. P. Williams, *J. Chem. Soc.*, 3192 (1953).
16. H. Taube, *Electron Transfer Reactions of Complex Ions in Solution*, Academic Press, Inc., New York, 1970; A. Haim, *Acc. Chem. Res.* **8**, 265 (1975); see Ref. 1.
17. D. F. Shriver, P. W. Atkins, and C. H. Langford, *Inorganic Chemistry*, W. H. Freeman and Co., New York, 1990, Chapt. 17.
18. G. J. Ferraudi, *Elements of Inorganic Photochemistry*, John Wiley & Sons, Inc., New York, 1988; A. W. Adamson and P. D. Fleischauer, eds., *Concepts of Inorganic Photochemistry*, John Wiley & Sons, Inc., New York, 1975, 439 pp.; C. R. Bock and E. A. Koerner von Gustorf, in J. N. Pitts, Jr., G. S. Hammond, and K. Gollnick, eds., *Advances in Photochemistry*, Vol. 10, John Wiley & Sons, Inc., New York, 1977, pp. 221–310; V. Balzani and V. Carositi, *Photochemistry of Coordination Compounds*, Academic Press, Ltd., London, 1970, 432 pp.
19. D. J. Cole-Hamilton and D. W. Bruce, in Ref. 9, Vol. 6, 1987, pp. 487–540.
20. L. DeCola and co-workers, *J. Am. Chem. Soc.* **110**, 7210 (1988).
21. A. Spencer, in Ref. 9, Vol. 6, 1987, Chapt. 66.
22. A. Spencer, in Ref. 9, Vol. 6, 1987, pp. 229–316; J. Halpern, *Inorg. Chim. Acta* **50**, 11 (1981); Ref. 11, Chapt. 28; see Ref. 17.
23. M. Strem, *The Strem Catalog, No. 8*, Strem, Chemical, Newburyport, Mass., 1978, p. 115.
24. W. S. Knowles, *Acc. Chem. Res.* **16**, 106 (1983); J. Halpern, *Pure Appl. Chem.* **55**, 99 (1983).
25. H. Mimoun, in Ref. 9, Vol. 6, 1987, pp. 317–410.
26. A. Spencer, in Ref. 9, Vol. 6, 1987, pp. 258–266.
27. A. Spencer, Ref. 9, Vol. 6, 1987, pp. 269, 275, 278, 292; W. A. R. Slegeir, R. S. Sapienza, and B. Easterling, *ACS Symp. Ser.* **152**, 325 (1981); R. M. Laine, R. G. Rinker, and P. C. Ford, *J. Am. Chem. Soc.* **99**, 252 (1977); P. C. Ford and A. Rokicki, *Adv. Organometal Chem.* **28**, 139 (1988).
28. D. W. Van Krevelen, *Chem. Ind.* **49**, 1396 (1971); F. C. A. A. Van Berkel and H. Grotjahn, *Appl. Polym. Symp.* **21**, 67 (1973); *Text. Prog.* **8**, 124, 156 (1976).
29. C. E. Carraher, Jr., C. U. Pittman, Jr., J. E. Sheats, M. Zeldin, and B. Currell, eds., *Inorganic and Metal-Containing Polymeric Materials*, Plenum Publishing Co., New York, 1991; see Chapt. I and earlier reviews cited therein.
30. R. D. Archer, V. J. Tramontano, V. O. Ochaya, P. V. West, and W. G. Cumming, in Ref. 29, pp. 161–171.
31. D. D. Chapman and E. R. Schmittou, Ref. 9, Vol. 6, pp. 95–132; T. H. James, ed., *The Theory of the Photographic Process*, 4th ed., Macmillan, New York, 1977.
32. J. Kosar, *Light Sensitive Systems: Chemistry and Applications of Nonsilver Halide Photographic Processes*, John Wiley & Sons, Inc., New York, 1966; E. Brinckman, G. Delzenne, A. Poot, and J. Willems, *Unconventional Imaging Processes*, Focal Press, London, 1978.
33. S. J. Lippard, ed., *Platinum, Gold and Other Metal Chemotherapeutic Agents*, *ACS Symp. Series 209*, American Chemical Society, Washington, D.C., 1983; H. E. Howard-Lock and C. J. L. Lock, in Ref. 9, Vol. 6, 1987, pp. 755–778.
34. P. J. Sadler, *Adv. Inorg. Chem.* **36**, 1 (1991).
35. C. J. Jones, in Ref. 9, Vol. 6, 1987, pp. 963–1009.
36. A. M. Pyle and J. K. Barton, *Progr. Inorg. Chem.* **38**, 413 (1987).
37. G. J. Hutchings, *Chem. in Britain*, 762 (1987); M. S. Spencer, *Crit. Rev. Appl. Chem.* **12**, 65 (1985).
38. M. J. Nicol, C. A. Fleming, and J. S. Preston, in Ref. 9, Vol. 6, 1987, pp. 779–842.
39. C. J. Jones, in Ref. 9, Vol. 6, 1987, pp. 881–963.

General References

- J. C. Bailar, Jr., ed., *Chemistry of Coordination Compounds*, *ACS Monograph 131*, Reinhold Publishing Corp., New York, 1956. Dated but very good on the historical and classical aspects of coordination chemistry.
- F. Basolo and R. C. Johnson, *Coordination Chemistry*, 2nd ed., Sci. Revs., Northwood, UK, 1986. Easy-to-read paperback on coordination chemistry suitable for anyone who has studied basic chemistry.
- F. Basolo and R. G. Pearson, *Mechanisms of Inorganic Reactions*, 2nd ed., John Wiley & Sons, Inc., New York, 1967. An excellent volume that stresses the reactions of complexes in solution; a background and a detailed theory section is included that is largely crystal field theory, but some advantages and disadvantages of molecular orbital theory are included.
- B. D. Berezin, V. G. Vopian, trans., *Coordination Compounds of Porphyrins and Phthalocyanines*, John Wiley & Sons, Chichester, UK, 1981. The synthesis, structure, and properties of the macrocycles and their coordination compounds are included.
- P. S. Braterman, *Reactions of Coordinated Ligands*, Plenum Press, New York; Vol. 1, 1986, includes over 1000 pages of carbon bonded ligand reactions, essentially organometallic and carbonyls; Vol. 2, 1989, discusses noncarbon bonded ligands.
- F. A. Cotton and G. Wilkinson, *Advanced Inorganic Chemistry*, 5th ed., Interscience Publishers, a division of John Wiley & Sons, Inc., New York, 1988.
- G. B. Kauffman, *Inorganic Coordination Compounds*, Heyden, London, 1981. An excellent volume on the history of coordination compounds prior to 1935.
- S. Kawaguchi, *Variety in Coordination Modes of Ligands in Metal Complexes*, Springer-Verlag, Berlin, 1988.
- L. F. Lindoy, *The Chemistry of Macrocyclic Ligand Complexes*, Cambridge University Press, Cambridge, UK, 1989.
- W. A. Nugent and J. M. Mayer, *Metal-Ligand Multiple Bonds: The Chemistry of Transition Metal Complexes Containing Oxo, Nitrido, Imido, Alkylidene, or Alkylidyne Ligands*, John Wiley & Sons, Inc., New York, 1988. Contains electronic and molecular structure, nmr, and ir spectroscopy, reactions, and catalysis.
- G. W. Parshall, *Homogeneous Catalysis: The Applications and Chemistry of Catalysis by Soluble Transition Metal Complexes*, John Wiley & Sons, Inc., New York, 1980, 240 pp. An excellent treatment of catalysis by coordination compounds.
- R. A. Sheldon and J. K. Kochi, *Metal-Catalyzed Oxidations of Organic Compounds*, Academic Press, Inc., New York, 1981. Mechanistic principles and synthetic methodology for homogeneous and heterogeneous metal-containing oxidation catalysts, many of which are coordination compounds, are discussed.
- R. van Eldik, ed., *Inorganic High Pressure Chemistry*, Elsevier, Amsterdam, The Netherlands, 1986. High pressure coordination kinetics including solvent exchange, octahedral and four-coordinate substitution, electron transfer, photochemical, and bioinorganics are discussed.
- R. G. Wilkins, *Kinetics and Mechanism of Reactions of Transition Metal Complexes*, 2nd ed., VCH, Weinheim, Germany, 1991. A critical and selected compilation of kinetics and mechanism data.
- G. Wilkinson, R. D. Gillard, and J. A. McCleverty, eds., *Comprehensive Coordination Chemistry*, Pergamon Press, Oxford, UK, 1987, 7 Vols., Vol. 1, theory and background; Vol. 2, ligands; and Vol. 6, applications.
- A. F. Williams, C. Floriani, and A. E. Merbach, *Perspectives in Coordination Chemistry*, VCH, Weinheim, Germany, 1992. Some state-of-the-art reviews included, July 1992.
- J. J. Zimlkowski, ed., *Coordination Chemistry and Catalysis*, World Scientific, Singapore, 1988. Compiles a very diverse collection of lectures by reknown leaders in this field.

Ronald D. Archer
University of Massachusetts, Amherst



COPOLYMERS

Synthetic polymers have become extremely important as materials over the past 50 years and have replaced other materials because they possess high strength-to-weight ratios, easy processability, and other desirable features. Used in applications previously dominated by metals, ceramics, and natural fibers, polymers make up much of the sales in the automotive, durables, and clothing markets. In these applications, polymers possess desired attributes, often at a much lower cost than the materials they replace. The emphasis in research has shifted from developing new synthetic macromolecules toward preparation of cost-effective multicomponent systems (ie, copolymers, polymer blends, and composites) rather than preparation of new and frequently more expensive homopolymers. These multicomponent systems can be "tuned" to achieve the desired properties (within limits, of course) much easier than through the total synthesis of new macromolecules.

This article reviews the preparation, properties, characterization, use, economic importance, and future of some of the important synthetic copolymers. Polymer blends (qv) and composites (qv) are also mentioned. Biocopolymers are not included.

Homopolymers and Copolymer Structures

Homopolymers are high molecular-weight molecules prepared by linking a large number of smaller molecules called monomers (eq. 1).



Macromolecules in which two or more different monomers (comonomers) are incorporated in the same polymer chain are copolymers (eq. 2).



Copolymers can be further described by specifying the number and distribution of monomer units within the copolymer molecule. Thus a polymer with a statistical placement of monomer units, eg, ABBABABAAB is called a random copolymer. An alternating copolymer consists of an alternating arrangement of the comonomers, ABABABABA. On the other hand, a block copolymer has a long segment of one monomer followed by a long segment of another monomer, AAAABBBB. A graft copolymer is a type of block copolymer in which a polymer chain (main chain or backbone) has chains (branched chains) of another polymer attached at intervals along the backbone. A network copolymer is a cross-linked or three-dimensional copolymer.

AAAAA	AAAAABAAAA
B	C
B	C
B	C
B	C
B	AAAAABAAAA

A polymer blend is a physical or mechanical blend (alloy) of two or more homopolymers or copolymers. Although a polymer blend is not a copolymer according to the above definition, it is mentioned here because of its commercial importance and the frequency with which blends are compared with chemically bonded copolymers. Another technologically significant material relative to the copolymer is the composite, a physical or mechanical combination of a polymer with some unlike material, eg, reinforcing materials such as carbon black, graphite fiber, and glass (see COMPOSITE MATERIALS).

These definitions represent only the most basic structures. They should apply to copolymers regardless of the mechanism of their formation, but their simplicity can be misleading, particularly in the case of step-growth copolymers. For example, an -(A-B)- polyamide, prepared from a diacid and a diamine, could be considered either a homopolymer or a copolymer. In this article the convention of Reference 1 has been adopted, according to which step-growth copolymers are only those materials that have used more than the minimum number of monomers to effect homopolymerization (eg, a copolyamide from a diacid and two diamines) or that have used prepolymers as monomers. This greatly reduces the number of step-growth materials that qualify as copolymers. Consequently, this article emphasizes chain-growth copolymers and copolymerization.

Nomenclature. There are myriad ways in which two or more monomers can be bonded to form copolymers. Consequently, a systematic nomenclature for copolymers is both a need and a challenge. The IUPAC source-based nomenclature for copolymers covers many cases (2,3). This method is based on homopolymer nomenclature that uses *poly* plus the name of the monomer, eg, poly(methyl methacrylate) and polystyrene. Copolymer names specify both monomers; the various arrangements of monomers are indicated by "infixes." Thus a copolymer composed of butadiene and styrene wherein the arrangement is unspecified is named poly(butadiene-*co*-styrene), poly(butadiene-*ran*-styrene) would indicate randomness, and an alternating copolymer of propylene oxide and carbon dioxide is named poly(propylene oxide-*alt*-carbon dioxide). Block and graft copolymers are distinguished by the infixes -*block*- and -*graft*- respectively, and the names are written poly A-*graft*-poly B. The first polymer segment (A) corresponds to the homopolymer or the copolymer that is the backbone; poly B names the branch.

Copolymers extend the number and range of available materials, enabling the polymer scientist to achieve combinations of material properties (eg, tensile strength, solubility, solvent resistance, low temperature flexibility, etc) unattainable from the simple constituent homopolymers. As a result, a large number of copolymers have become commercially important. Table 1 lists some of them.

Table 1. Some Commercially Important Copolymers

Comonomers	Nomenclature	CAS Registry Number	CAS Indexing	Generic name, trade name, and or abbreviation
butadiene–styrene	poly(butadiene- <i>co</i> -styrene)	[9003-55-8]	benzene, ethenyl-, polymer with 1,3 butadiene	synthetic rubber; GRS, SBR
ethylene–propylene	poly(ethylene- <i>co</i> -propylene)	[9010-79-1]	ethene, polymer with 1-propene	EPM, EPR rubber
ethylene–propylene–diene	poly(ethylene- <i>co</i> -propylene- <i>co</i> -5-ethylidene-2 norbornene)	[25038-36-2]	ethene, polymer with 1-propene and 5-ethylidene-2-norbornene	EPDM rubber
butadiene–acrylonitrile	poly(butadiene- <i>co</i> -acrylonitrile)	[9003-18-3]	2-propenenitrile, polymer with	NBR rubber

styrene–butadiene–styrene (triblock)	polystyrene- <i>block</i> -polybutadiene- <i>block</i> -polystyrene	[9003-55-8]	1,3-butadiene	SBS thermoplastic rubber;
isobutylene–isoprene	poly(isobutylene- <i>co</i> -isoprene)	[9010-85-9]	benzene, ethenyl-polymer with 1,3 butadiene	Kraton
			2-methyl- 1-propene, polymer with 2-methyl- 1,3-butadiene	butyl rubber, GR-1
vinyl chloride–vinyl acetate	poly(vinyl chloride- <i>co</i> -vinyl acetate)	[9003-22-9]	acetic acid vinyl ester, polymer with chloroethene	Vynlite flooring, Tygon tubing, coatings
vinyl chloride–acrylonitrile	poly(vinyl chloride- <i>co</i> -acrylonitrile)	[9003-00-3]	2-propenenitrile, polymer with chloroethone	Dynel fibers
vinyl choride–vinylidene chloride	poly(vinyl chloride- <i>co</i> -vinylidene chloride)	[9011-06-7]	ethene, 1,1-dichloro-, polymer with chloroethene	Saran packaging, fibers
acrylonitrile–vinyl acetate	poly(acrylonitrile- <i>co</i> -vinyl acetate)	[24980-62-9]	2-propenenitrile, polymer with ethenyl acetate	Orlon, Acrilon acrylic fibers
acid–glycol–diisocyanate (step growth multiblock) ^a	poly(acid- <i>co</i> -glycol- <i>co</i> -diisocyanate- <i>co</i> -diamine)		the acid, polymer with the glycol; the diisocyanate, and the diamine	Spandex fibers
acrylonitrile–butadiene–styrene ^b	polybutadiene- <i>graft</i> -poly(styrene- <i>co</i> -acrylonitrile) + poly(styrene- <i>co</i> -acrylonitrile)	[9003-56-9]	2-propenenitrile, polymer with 1,3-butadiene and ethenylbenzene	ABS plastics
butadiene–styrene ^b	polybutadiene- <i>graft</i> -polystyrene + polybutadiene + polystyrene	[9003-55-8]	benzene, ethenyl-, polymer with 1,3-butadiene	IPS impact styrene plastics
styrene–acrylonitrile	poly(styrene- <i>co</i> -acrylonitrile)	[9003-54-7]	2-propenenitrile, polymer with ethenylbenzene	SAN plastics
styrene–maleic anhydride	poly(styrene- <i>alt</i> -maleic anhydride)	[9011-13-6]	2,5, furandione, polymer with ethenylbenzene	SMA resins

^a See Equation 30.
^b Graft copolymer and polymer blend.

Copolymerization Reactions

The mutual polymerization of two or more monomers is called copolymerization. This topic has been comprehensively reviewed (4,5). Monomers frequently show a different reactivity toward copolymerization than toward homopolymerization. In fact, some monomers that can be homopolymerized only with great difficulty, can be readily copolymerized. One such monomer is maleic anhydride. It is rather inert to free-radical homopolymerization yet can be copolymerized conveniently with styrene and many other monomers under free-radical conditions.

Chain-Growth Copolymerization Theory. The theory of chain-growth (eg, radical, anionic, etc) copolymerization has received more attention than that of step-growth or other copolymerizations. In the case of chain-growth copolymerization, growing polymer chains must choose between more than one monomer. Such a choice or relative reactivity has been quantitatively treated by the reactivity ratio (6,7) and the *Q-e* schemes (8).
Reactivity Ratio Scheme. The composition of a copolymer at any point in time depends on the relative rates that each monomer can add to a chain end. If it is assumed that the chemical reactivity of a propagating chain depends only on the terminal unit and is not affected by any penultimate units, then four possible propagation steps in the copolymerization of two monomers, M₁ and M₂, with two growing chain ends, M₁^{*} and M₂^{*}, can be written as follows:



The reactivity ratios *r*₁ and *r*₂ are defined as follows:

$$r_1 = \frac{k_{11}}{k_{12}}$$

(7)

$$r_2 = \frac{k_{22}}{k_{21}}$$

(8)

The ratio *r*₁ describes the relative reactivity of polymer chain M₁^{*} toward monomer M₁ and monomer M₂. Likewise, *r*₂ describes the relative reactivity of polymer chain M₂^{*} toward M₂ and M₁. With a steady-state assumption, the copolymerization equation can be derived from the propagation steps in equations 3–6.

$$\frac{d[M_1]}{d[M_2]} = \frac{[M_1]}{[M_2]} \frac{r_1[M_1] + [M_2]}{r_2[M_1] + [M_2]}$$

(9)

This equation relates the (instantaneous) copolymer composition with the monomer feed of M₁ and M₂. Values for *r*₁ and *r*₂ are usually determined by graphical methods (9,10). Today, with the prevalence of powerful desktop computers, numerical minimization methods are often used (11–14).

The combined values of the reactivity ratios (ie, *r*₁*r*₂) yield important clues regarding the composition of the copolymer. Reactivity ratio combinations are generally classified in five categories (15) designated Types I–V in the discussion.*Type I*

$$r_1 \approx r_2 \approx 1, \; (r_1r_2 = 1)$$

(10)

Both r_1 and r_2 are approximately unity in an ideal copolymerization. In this case, $k_{11} \approx k_{12}$ and $k_{22} \approx k_{21}$ and the growing chains show little preference for either monomer, M_1 or M_2 . This copolymerization is essentially random, and the composition of the copolymer is approximately equal to that of the monomer feed at all feed compositions (Fig. 1). *Types II*

$$r_1 \text{ and } r_2 < 1, (r_1 r_2 < 1)$$

(11)

In the Type II case, the copolymerization tends toward an alternating arrangement of monomer units. Curve II of Figure 1 shows an example of an alternating copolymer that has an azeotropic copolymer composition, ie, a copolymer composition equal to the monomer feed at a single monomer feed composition. This case is analogous to a constant boiling mixture in vapor–liquid equilibria. *Type III*

$$r_1 \text{ and } r_2 > 1, (r_1 r_2 > 1)$$

(12)

When both r_1 and r_2 are greater than 1, the copolymer becomes blocky in nature. When r_1 and r_2 are much greater than 1, homopolymers of each monomer are formed and the resulting material is usually a polymer blend of some copolymer and homopolymer. *Type IV*

$$r_1 \gg 1 \text{ and } r_2 \ll 1$$

(13)

In the Type IV case, copolymerization is very difficult to achieve, especially when M_1 is high (curve IV of Fig. 1). Monomers with such a combination of reactivity ratios and monomer charge tend to give homopolymer. However, when M_1 has been nearly exhausted, copolymerization becomes more favorable. *Type V*

$$r_1 \approx r_2 \approx 0, (r_1 r_2 \approx 0)$$

(14)

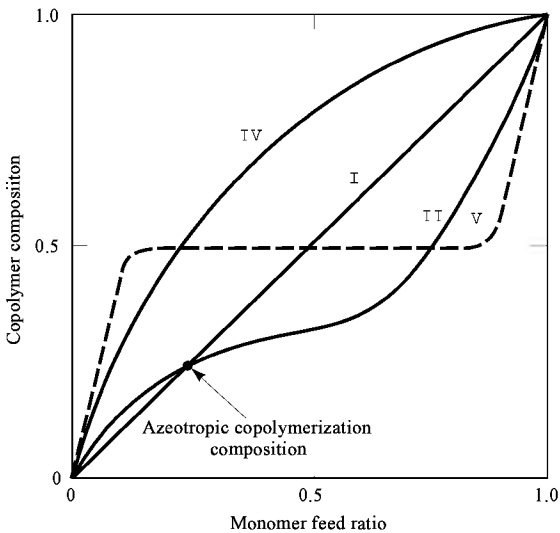


Fig. 1. Copolymer composition as a function of monomer feed ratio for various reactivity ratio combinations, designated I–V and explained in the text (16).

Type V represents the case in which the copolymerization is strongly alternating (see Fig. 1). Here, each growing chain reacts exclusively with the unlike monomer.

Table 2 shows characteristic reactivity ratios for selected free-radical, ionic, and coordination copolymerizations. The reactivity ratios predict only tendencies; some copolymerization, and hence some modification of physical properties, can occur even if r_1 and r_2 are somewhat unfavorable. For example, despite their dissimilar reactivity ratios, ethylene and propylene can be copolymerized to a useful elastomeric product by adjusting the monomer feed or by using a catalyst that increases the reactivity of propylene relative to ethylene.

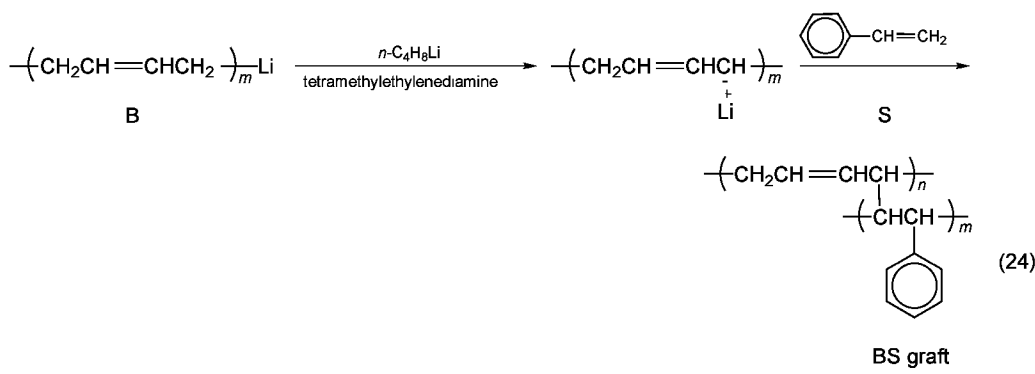
Table 2. Characteristic Reactivity Ratios

M ₁	M ₂	r ₁	r ₂	Synthetic method	Reference
styrene	1,3-butadiene	0.82	1.38	free radical	17
styrene	vinyl acetate	56.0	0.01	free radical	18
styrene	maleic anhydride	0.04	0.0	free radical	19
acrylonitrile	vinyl acetate	5.5	0.9	free radical	20
methyl methacrylate	vinyl chloride	0.84	0.99	free radical	21
ethylene	propylene	17.95	0.06	coordination, 25°C	22
ethylene	vinyl acetate	1.2	1.1	free radical, 160°C	23
styrene	1,3-butadiene	0.1	12.5	anionic, ^a 25°C	24
styrene	1,3-butadiene	0.11	1.78	anionic, ^b 30°C	25
styrene	1,3-butadiene	0.01	1.30	cationic, 43°C	26

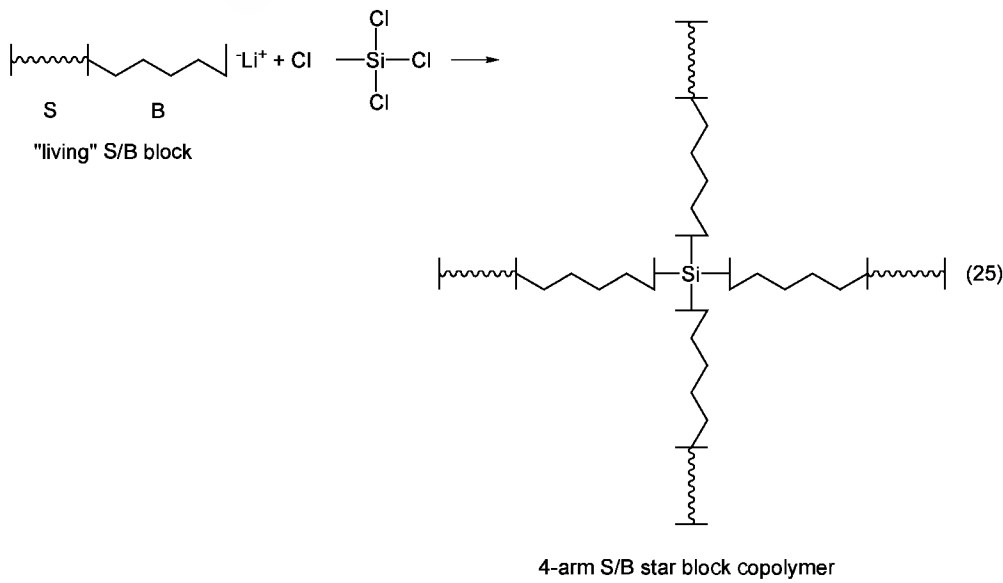
^a Alkylolithium initiation in toluene.
^b Alkylolithium initiation in diethyl ether.

The Q-e Scheme. The magnitude of r_1 and r_2 can frequently be correlated with structural effects, such as polar and resonance factors. For example, in the free-radical polymerization of vinyl acetate with styrene, both styrene and vinyl acetate radicals preferentially add styrene because of the formation of the resonance stabilized polystyrene radical.

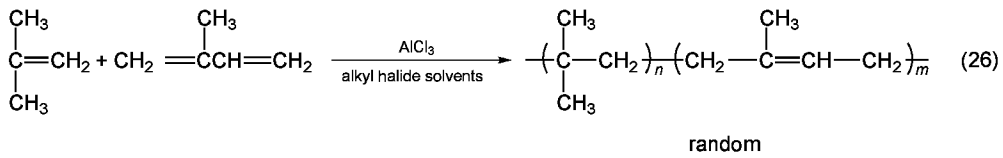
Alfrey and Price proposed a means of predicting monomer reactivity in copolymerization from two parameters, Q (a measure of resonance) and e (a measure of polar effects) (8). These parameters have been related to the reactivity ratios by equations 15–17.

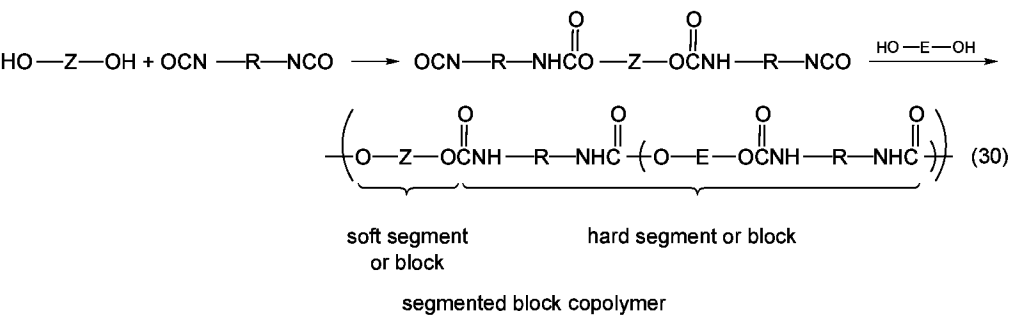


Anionic polymerization techniques are one of many ways to synthesize a special class of block copolymers, referred to as star block copolymers (eq. 25) (33). Specifically, a "living" SB block is coupled with a silyl halide coupling agent. The term living polymerization refers to a chain polymerization that proceeds in the absence of termination or transfer reactions.

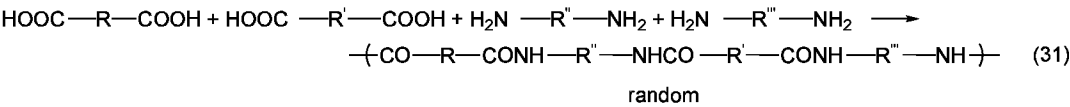


Cationic Copolymerization. Illustrative cationic copolymerizations are shown in equations 26 and 27 (34,35).

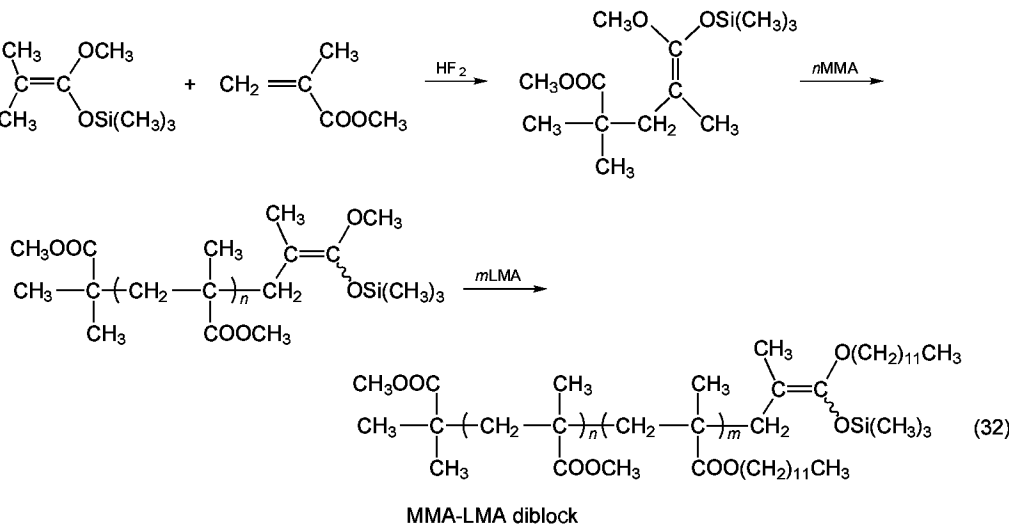




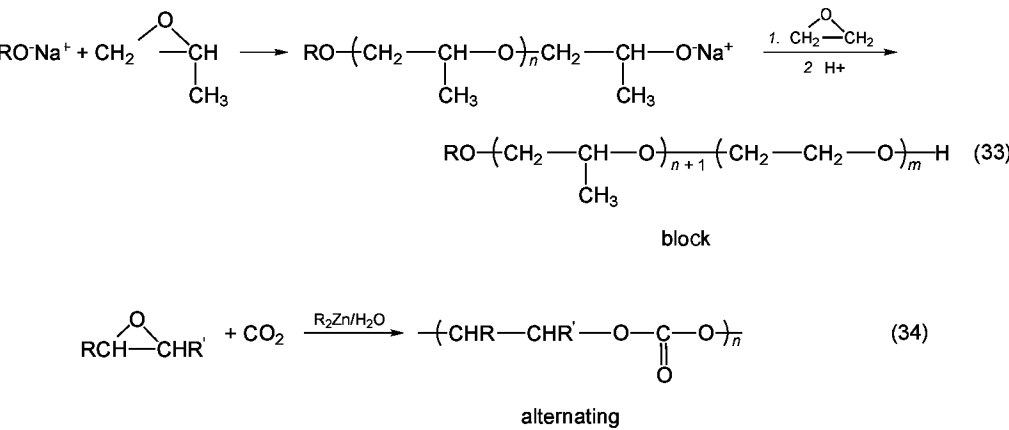
Random copolymers can be synthesized by step-growth copolymerization in an equimolar mixture of four monomers (1).



Group-Transfer Polymerization. Living polymerization of acrylic monomers has been carried out using ketene silyl acetals as initiators. This chemistry can be used to make random, block, or graft copolymers of polar monomers. The following scheme demonstrates the synthesis of a methyl methacrylate–lauryl methacrylate (MMA–LMA) AB block copolymer (38). LMA is $\text{CH}_2=\text{C}(\text{CH}_3)\text{COO}(\text{CH}_2)_{11}\text{CH}_3$.



Ring-Opening Polymerization. Examples of the formation of copolymers by ring-opening reactions are shown in equations 33 and 34 (39,40).



Ring-Opening Metathesis Polymerization. Several new titanacyclobutanes have been shown to initiate living ring-opening metathesis polymerization (ROMP) systems. These have been used to make diblock and triblock copolymers of norbornene [498-66-8] (N) and its derivatives (eg, dicyclopentadiene [77-73-6] (D)) (Fig. 2) (41).

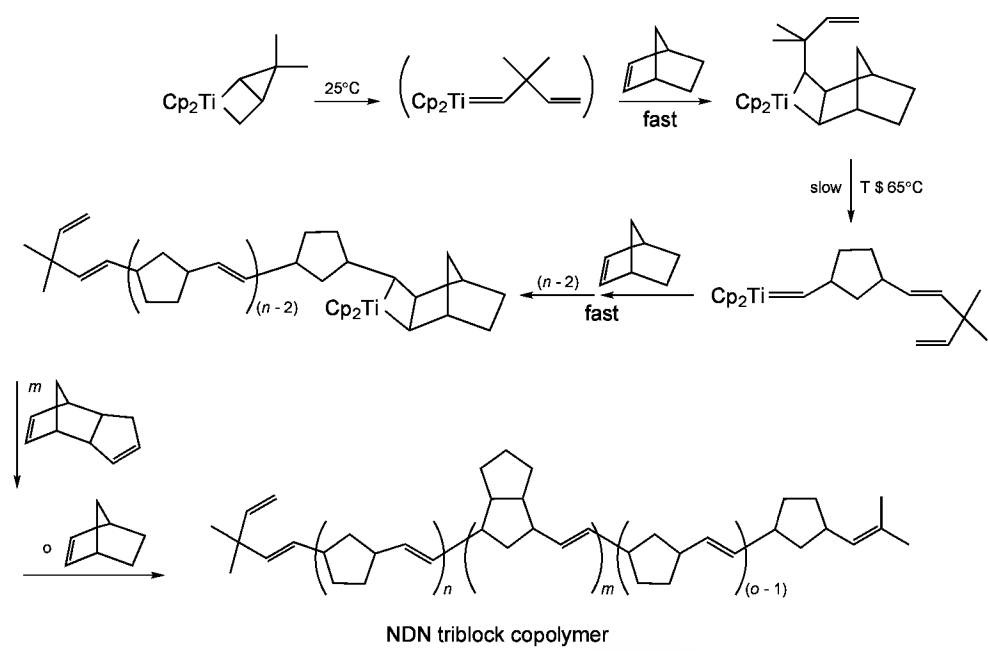
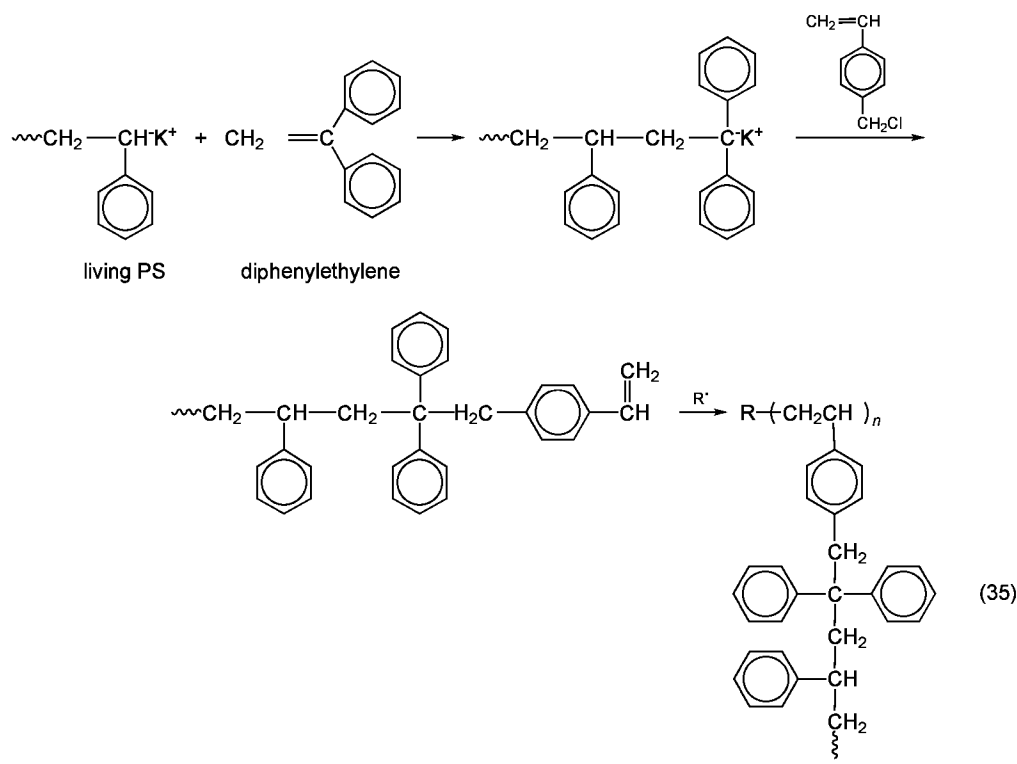
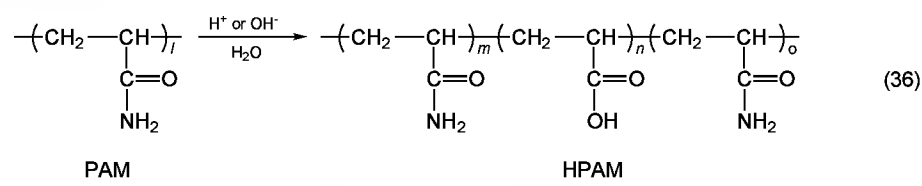


Fig. 2. Ring-opening metathesis polymerization. Cp — cyclopentadienyl.

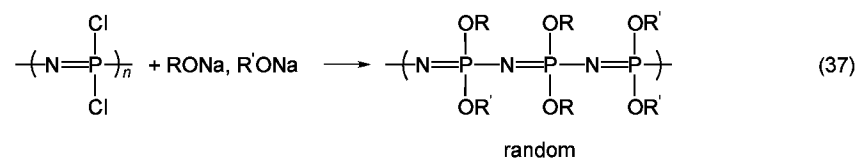
Telechelic Polymers and Macromonomers. The synthesis of many block and graft copolymers has been achieved through the use of macromonomers and telechelic polymers. A telechelic polymer is one containing one or more functional end groups that have the capacity for selective reaction to form bonds with another molecule. Telechelic polymers are often also called macromonomers because it is possible to use them as monomers in step- and chain-growth polymerizations (42,43). Equation 35 is an example, where R^{*} represents a free-radical initiator and the final product is a *p*-methylstyrene–styrene graft copolymer.



Postpolymerization Reactions. Copolymers can also be formed by postpolymerization reactions on polymers. A well-known example is the partial hydrolysis of polyacrylamide (PAM) to hydrolyzed polyacrylamide (HPAM). The product becomes a random copolymer of acrylamide and acrylic acid (44) (see ACRYLAMIDE POLYMERS).



A random copolymer can be formed by postpolymerization reaction with a mixture of reagents, eg, in the case of polyphosphazenes (eq. 37) (45,46).



Effect of Monomer Unit Arrangement on Physical Properties

Random Arrangement. The primary incentive for preparing copolymers is to attain certain properties in the products. The effect of random copolymerization on polymer properties is easily shown by differences in polymer crystallinity, melting point (T_m), glass-transition temperature (T_g), and solubility between a copolymer and the corresponding homopolymers. Because random comonomer enchainment tends to reduce symmetry and modify intermolecular forces, it is not surprising that random copolymers have different melting behavior than the corresponding constituent crystalline homopolymers. Figure 3 shows the variation in T_m of two copolymers. The melting point of the adipamide copolymer varies smoothly over the entire composition range. On the other hand, the sebacamide copolymer shows a eutectic point.

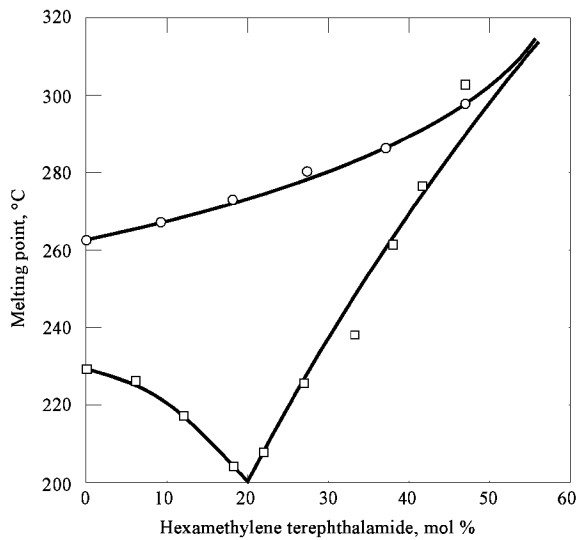


Fig. 3. Melting point versus composition for polyamide copolymers. The comonomers employed are adipic acid, forming adipamide copolymers (○), and sebacic acid, forming sebacamide copolymer (□) (47).

The glass-transition temperature in amorphous polymers is also sensitive to copolymerization. Generally, T_g of a random copolymer falls between the glass-transition temperatures of the respective homopolymers. For example, T_g for solution-polymerized polybutadiene is -96°C ; that for solution-polymerized polystyrene is $+100^{\circ}\text{C}$. A commercial solution random copolymer of butadiene and styrene (Firestone's Stereon) shows an intermediate T_g of -76°C (48). The glass-transition temperature of the random copolymer can sometimes be related simply as follows:

$$1/T_g = W_1/T_{g1} + W_2/T_{g2} \tag{38}$$

T_{g1} and T_{g2} are the glass-transition temperatures in K of the homopolymers; W_1 and W_2 are the weight fractions of the comonomers (49). Because the glass-transition temperature is directly related to many other material properties, changes in T_g by copolymerization cause changes in other properties too. Polymer properties that depend on the glass-transition temperature include physical state, rate of thermal expansion, thermal properties, torsional modulus, refractive index, dissipation factor, brittle impact resistance, flow and heat distortion properties, and minimum film-forming temperature of polymer latex (50).

The solubility of random copolymers of monomers whose homopolymers are noncrystalline also varies quite regularly as the relative amounts of the comonomers are changed. The solubility of random copolymers is often low in solvents for the respective homopolymers but high in solvent pairs (51).

Another important feature of some random copolymers is the ability to achieve miscibility in either a homopolymer or a second random copolymer. This "copolymer effect" has been shown empirically for quite some time, eg, PVC is miscible with random copolymers of ethylene and vinyl acetate (52). Such systems are effective because repulsions between the dissimilar segments in the copolymer are enough to overcome the repulsions between these segments and those of the second component in the mixture. In other words, in the above example, the ethylene units "hate" vinyl acetate units more than either of them "hate" PVC. Thus there is a net negative interaction energy and the two materials are miscible (53).

Block (Linear) Arrangement. By far the most interesting block copolymers are those in which there is little or no mixing of the block phases. Such heterogeneous block copolymers tend to show the properties of the components rather than the averaging of homopolymer properties. Heterogeneous copolymers also tend to be soluble in a wide variety of solvents. In fact, they may act like surfactants if the respective blocks have widely different solubilities. Because the domain sizes in block copolymers are often much smaller than those of polymer blends, the block copolymers tend to be more optically transparent than blends. The surfactant behavior may be exploited by using heterogeneous block copolymers to compatibilize immiscible homopolymers. AB diblocks or ABA triblocks can be added to a blend of homopolymer A and homopolymer B to enlarge the miscible region (54). In most commercial systems, miscibility is not required; in fact, two-phase systems are desired. In such cases, block copolymers are added to two immiscible homopolymers to control morphology and domain size. It is believed that the block copolymer resides at the interface between homopolymer A and homopolymer B, acting as "polymeric soap" or compatibilizer. The overall effect is to reduce interfacial tension and "connect" the two phases. This generally results in property enhancements (55).

Not only do block copolymers have properties different from those of homopolymers, random copolymers, and polymer blends but the properties of block copolymers themselves differ, depending on the length and chemical composition of the blocks. This results in interesting morphologies and structures. When critical molecular weights are attained, microphase separation occurs. In styrene-butadiene diblock copolymers, for example, morphology is greatly affected by the chain length of the block. Three basic morphological units are observed: spheres, cylinders, and lamellae (56). The effect of composition on these structures is depicted in Figure 4.

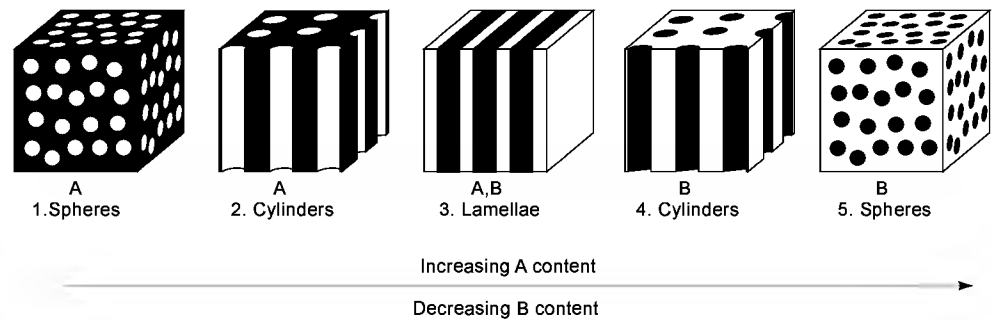


Fig. 4. Structure of an AB block copolymer with, from left to right, an increasing chain length of the A block; 1, spheres of A in a continuous matrix of B; 2, cylinders of A in a continuous matrix of B; 3, lamellae, 4, cylinders of B in a continuous matrix of A; and 5, spheres of B in a continuous matrix of A (56).

Whereas random copolymers exhibit one T_g described by equation 38, block copolymers, because of this microphase separation, exhibit two glass-transition temperatures. The T_g of each block is close to, if not the same as, the homopolymer from which it was formed. Polymer properties are also affected by the arrangement of the blocks. This is shown for high styrene-containing or high molecular-weight styrene resins of various block arrangements in Table 3.

Table 3. Properties of High Styrene–Butadiene Block Copolymers^a

Samples	S B mole ratios ^b	Elongation at break, % ^c	Light transmittance, %	Charpy impact strength, J m ^c	Heat distortion temperature, °C
S	100 0	7	90	20.6	97
SB	60 40	2	62	24.5	83
SBS	60 40	70	71	43.2	67
BSB	60 40	2	51	23.5	78
SBSB	60 40	23	70	30.4	55
commercial HIPS ^d		30	0	68.6	75
polymer blend	60 40	2	0	4.9	

^a Ref. 57.
^b Charged to reactor.
^c To convert J m to ft-lb in., divide by 53.38.
^d HIPS — high impact strength polystyrene, a graft copolymer and blend with polybutadiene.

Block (Star) Arrangement. The known star polymers, like their linear counterparts, exhibit microphase separation. In general, they exhibit higher viscosities in the melt than their analogous linear materials. Their rheological behavior is reminiscent of network materials rather than linear block copolymers (58). Although they have been used as compatibilizers in polymer blends, they are not as effective at property enhancements as linear diblocks (59).

Graft Arrangement. Graft copolymers typically exhibit some microphase separation, especially in cases in which the backbone polymer has different solubility compared with the chains grafted onto it. The physical properties lie between those of a heterogeneous diblock or triblock copolymer and a random copolymer. When the components of graft copolymers are more soluble, more homogeneity exists and a single-phase morphology is observed. In this case, the physical properties of these materials are usually intermediate between those of the respective homopolymers. Homogeneous graft copolymers can result from low molecular-weight segments or from specific interactions, such as hydrogen bonding.

Because graft copolymers are much "easier" to obtain synthetically than heterogeneous diblock or triblock copolymers, they have also been used as compatibilizers in polymer blends. Theoretically, they are not as efficient as the diblocks (60), but they are successfully and economically used in a number of commercial systems (61).

Random Copolymers

Many random copolymers have found commercial use as elastomers and plastics. For example, SBR (62), poly(butadiene-*co*-styrene) [9003-55-8], has become the largest volume synthetic rubber. It can be prepared in emulsion by use of free-radical initiators, such as K₂S₂O₈ or Fe²⁺ ROOH (eq. 18), or in solution by use of alkyl lithium initiators. Emulsion SBR copolymers are produced under trade names by such companies as American Synthetic Rubber (ASPC), Amtek, B. F. Goodrich (Ameripool), and Goodyear (Plioflex); solution SBR is manufactured by Firestone (Stereon). The total U.S. production of SBR in 1990 was 581,000 t (63).

SBR is a low T_g copolymer that is extendable with oil, reinforceable with carbon black, and vulcanizable with sulfur and an accelerator. Many of the physical properties of SBR vulcanizates (eg, tensile strength, resilience, hot tear strength, hysteresis, etc) are somewhat inferior to those of natural rubber (NR). However, its low cost, cleanliness, aging properties, and abrasion resistance make it an attractive rubber for use in passenger tires. Although most of the SBR produced in the United States has been used in tires, use in nonautomotive applications has been on the rise. For example, SBR use in wire and cable materials, mechanical goods, footwear, and foam products has increased.

Poly(butadiene-*co*-acrylonitrile) [9008-18-3], NBR (64), is another commercially significant random copolymer. This rubber is manufactured by free-radical emulsion polymerization. Important producers include Copolymer Rubber and Chemical (Nysyn), B. F. Goodrich (Hycar), Goodyear (Chemigum), and Uniroyal (Paracril). The total U.S. production of nitrile rubber (NBR) in 1990 was 95.6 t (65). The most important property of NBR rubber is its oil resistance. It is used in oil well parts, fuels, oil, and solvents (64) (see ELASTOMERS, SYNTHETIC NITRILE RUBBER).

Highly saturated nitrile elastomers (HSN) have become available. These rubbers are prepared by (nearly complete) hydrogenation of the nitrile rubber copolymer. The resulting product has better heat and oxidation resistance than conventional nitrile rubber but still retains some double bonds for vulcanization. Trade names for HSN are Zetpol (Nippon Zeon), Therbar (Bayer), and Tormac (Polysar). HSN has been used, and is being developed, for oil field chemical, automotive, power station, aerospace, military, and industrial applications (66).

Another important class of random copolymers are the ethylene–propylene elastomers. The saturated hydrocarbon backbone provides good ozone resistance and better weathering and aging characteristics than diene rubbers. Even though these materials are synthesized by Ziegler–Natta catalysts, they do not exhibit much stereospecificity. There are basically two types of these copolymers, ethylene–propylene (EPM) and ethylene–propylene–diene (EPDM) materials (67). EPM materials are completely saturated; if vulcanization is required, it is carried out by means of free-radical generators (ie, organic peroxides). These copolymers are typically polymerized in an organic solvent or by using the liquid phase of the monomer itself. Copolymers containing from 40 to 60 mol % of ethylene are most desirable; more ethylene leads to crystalline block structures. EPDM materials contain 3 to 10 wt % of one of the following nonconjugated dienes: ethylidene norbornene (ENB); 1,4-hexadiene (HD); or dicyclopentadiene (DCPD). Similarly, these random terpolymers are made in a solution process where the solvent is organic or the monomer itself is in the liquid phase. Suspension processes have also been developed. Because these materials contain unsaturation, they are easily vulcanized with either common rubber accelerators and sulfur or with organic peroxides. Important producers of EP elastomers include Copolymer Rubber and Chemical (EP syn), Du Pont (Nordel), Exxon Chemical (Vistalon), Polysar Inc. (Polysar EPM and EPDM), and Uniroyal (Royalene). The U.S. production of EPM and EPDM in 1990 was 295,000 t (68).

EPDM elastomers are low specific gravity rubbers with good vulcanizate properties (eg, tensile strength and elongation) and good weathering resistance. They are exceptional at accepting large amounts of oil, carbon black, or fillers while still maintaining these properties. Thus they have become important in automotive uses (radiator and heater hoses, weather stripping, and door and window seals), tires, and tubes. Other industrial uses include single-ply roofing, wire–cable insulation, other hoses, and appliance parts. One new and growing EPM–EPDM application is in the use of thermoplastic elastomers (TPE). EP elastomers have been used in polymer blends of polypropylene or polyethylene; they can be used either with or without cross-linking. These blends are attractive because a wide range of properties is available by varying blend composition, type of elastomer used, and processing conditions.

SAN resins are random, noncrystalline copolymers of styrene and acrylonitrile (69). These materials are manufactured by emulsion, suspension, or continuous mass polymerization. Polymer properties are adjusted by varying molecular weight and styrene-acrylonitrile ratio. Significant producers are Dow (Tyrl) and Monsanto (Lustran SANZI) (70). SAN copolymers are typically tougher and stronger than polystyrene. The presence of acrylonitrile imparts high heat deflection and chemical resistance to the copolymer. The important markets for SAN employ its excellent transparency, chemical resistance to common foods and detergents, and heat resistance sufficient to withstand dishwashers. Thus SAN is used primarily in housewares (tumblers, salad bowls, serving trays, etc) and appliances (refrigerator drawers, vacuum cleaner parts, washing machine detergent dispensers). Other important markets for SAN include automotive uses (dashboard components), furniture (chair backs, seats, lamps), medical disposables (parts for kidney dialysis equipment), packaging (containers, closures, dispensing parts), and alloys (blends with ABS and PVC (69)). In 1990, the total U.S. consumption of SAN was 60,000 t (71) (see ACRYLONITRILE POLYMERS).

Random copolymers of vinyl chloride and other monomers are important commercially. Most of these materials are produced by suspension or emulsion polymerization using free-radical initiators. Important producers for vinyl chloride–vinylidene chloride copolymers include Borden, Inc. and Dow. These copolymers are used in specialized coatings applications because of their enhanced solubility and as extender resins in plastisols where rapid fusion is required (72). Another important class of materials are the vinyl chloride–vinyl acetate copolymers. Principal producers include Borden Chemicals & Plastics, B. F. Goodrich Chemical, and Union Carbide. The copolymerization of vinyl chloride with vinyl acetate yields a material with improved processability compared with vinyl chloride homopolymer. However, the physical and chemical properties of the copolymers are different from those of the homopolymer PVC. Generally, as the vinyl acetate content increases, the resin solubility in ketone and ester solvents and its susceptibility to chemical attack increase, the resin viscosity and heat distortion temperature decrease, and the tensile strength and flexibility increase slightly.

Poly(vinyl chloride-*co*-vinyl acetate) [9003-22-9] has found application in flooring, phonograph records, protective coatings, fibers, and some films and sheeting. Because of their low viscosity and good processability, such copolymers constitute the bulk of the vinyl tile market. The total production of PVC copolymers in 1989 was 113,500 t (73) (see VINYL POLYMERS).

Poly(ethylene-*co*-vinyl alcohol) is made by saponification of ethylene–vinyl acetate copolymers. The properties of these materials depend on the amount of vinyl alcohol present in the copolymer. High vinyl alcohol content results in more hydrophilic materials possessing higher densities, stiffness, and moduli. They are used commercially as barrier resins for packaging. Important producers include Du Pont and EVALCA (74) (see BARRIER POLYMERS).

Fluoroelastomers are copolymers based on two or more of the following monomers: vinylidene fluoride (VF₂), hexafluoropropylene (HFP), chlorotrifluoroethylene (CTE), tetrafluoroethylene (TFE), and perfluoromethyl vinyl ether (FVE). They are made by a high pressure, free-radical initiated emulsion polymerization process. These materials are known for their chemical inertness and excellent thermal stability. Some commercial fluoroelastomers are peroxide curable and can be processed on typical rubber-processing equipment (75). These materials exhibit good tensile strengths and compression sets in addition to high heat resistance and low permeability. Du Pont produces VF₂–HFP copolymer (Viton A and E), VF₂–HFP–TFE terpolymer (Viton B), VF₂–TFE–FVE terpolymer (Viton G), and FVE–TFE copolymer (Kalrez). 3M produces VF₂–HFP copolymer (Fluorel FC), VF₂–HFP–TFE terpolymer (Fluorel FT), and VF₂–CTE copolymer (Kel F) (see ELASTOMERS, SYNTHETIC FLUOROCARBON ELASTOMERS; FLUORINE COMPOUNDS, ORGANIC).

The excellent properties of these fluoroelastomers come with a high price tag. Kalrez, for example, is extremely expensive (\$33–44 kg). These materials are used in automotive applications (seals, gaskets, fuel hose lines, engine parts, etc), where they can withstand under-hood temperatures. They are also used in equipment for oil and gas production and chemical processing. U.S. consumption in 1988 was 3100 t (76).

Alternating Copolymers

Poly(styrene-*alt*-maleic anhydride) [9011-13-6] is a classic and commercial example of an alternating copolymer (77). This material is manufactured by free-radical bulk, solution, or emulsion copolymerization. Important producers are ARCO (SMA) and Monsanto (Lytron). Such copolymer resins are brittle and insoluble in most solvents. But they are soluble in alkaline solution and react with water to give acids, with alcohols to give esters, and with amines to give amides. They can be converted to insoluble, infusible thermosets by heating with diamines or glycols. These resins and their derivatives are seldom used alone but are used as dispersants (to increase the pigment concentration) and floor polishes (to act as emulsifiers and protective colloids). Maleic anhydride can also form alternating copolymers with various olefins (EMA Resin, Monsanto) and vinyl ethers (Gantrez An-Resin, General Aniline and Film Corp.).

Maruzen Oil Co. has developed various Ziegler-Natta catalysts that can produce poly(butadiene-*alt*-propylene) (PBR) (78). PBR shows tack (self-adhesion) and green (unvulcanized) dynamic properties superior to those of BR and EPDM. Carbon black-loaded vulcanizates can be compounded to give high strength and elongation at break (79,80). PBR can also be covulcanized with SBR, BR, and EPDM.

Block Copolymers

Block copolymers have become commercially valuable commodities because of their unique structure–property relationships. They are best described in terms of their applications such as thermoplastic elastomers (TPE), elastomeric fibers, toughened thermoplastic resins, compatibilizers, surfactants, and adhesives (see ELASTOMERS, SYNTHETIC THERMOPLASTIC).

A thermoplastic elastomer is a material that combines the processability of a thermoplastic with the performance of a thermoset rubber. A thermoplastic elastomer (81) results when block copolymers have an ABA, (AB)_{*n*}X, or -(AB)-_n but not an AB diblock arrangement of A (thermoplastic) and B (rubbery) blocks. The hard A blocks may be glassy (eg, polystyrene) or crystalline (eg, polyester, polyurethane); the soft B blocks must be elastomeric (eg, polybutadiene, polyisoprene, polyether). When the hard segments are incompatible with the soft segments, the domains or regions of hard blocks act as reinforcing physical cross-links for the rubbery matrix. In contrast to chemically cross-linked rubbers, the physical network is thermally reversible. When the polymer is heated above the *T_g* (or *T_m*) of the hard block, the hard blocks soften and allow the rubber to flow and to be processed as a thermoplastic. Table 4 shows some commercially important TPE block copolymers and their producers.

Table 4. Commercially Important Block Copolymer TPE^a

Comonomer	Block arrangement	Trade name	Producer
styrene–diene (hydrogenated styrene–diene)	ABA, (AB) _{<i>n</i>} X	Kraton	Shell
		Vector	Dexco
urethane–ester (ether)	-(AB)-_n	Estane	B. F. Goodrich
		Texin	Mobay
ester–ether	-(AB)-_n	Hytrel	Du Pont

^a Ref. 82.

The manufacture of block copolymer TPE depends on the type and arrangement of the blocks. For example, butadiene–styrene ABA, (AB)_{*n*}X block copolymers are conveniently prepared by alkylolithium initiated anionic polymerization. Thermoplastic -(AB)-_n polyurethanes are synthesized by step-growth addition copolymerization of dihydroxy compounds such as polytetramethylene ether glycol and toluene diisocyanate. The copolyester–ether (AB) -_n copolymers are produced by the polycondensation of dicarboxylic acids (eg, terephthalic acid) with glycols or polyether glycols.

The physical properties of block copolymer TPE also depend on the type and arrangement of the blocks. Table 5 compares the property advantages of various block copolymer thermoplastic elastomers.

Table 5. Property Advantages of Various Block Copolymer TPE^{a, b}

Property	Styrene–diene	Hydrogenated styrene–diene	Ester–ether	Urethane–ester
tensile			+	+
recovery	+	+		
upper use temperature			+	+
lower use temperature	+	+		
aging stability		+		
acid–base resistance	+	+		
oil resistance			+	+
electrical	+	+		
abrasion resistance				+
melt processability			+	+
cost	+			

^a Ref. 83.
^b A designation of + indicates a performance strong point.

The properties and prices of the various block copolymer TPE greatly affect their markets. For example, the low cost butadiene–styrene block copolymers have found utility in footwear (sneakers, tennis shoes), injection-molded or extruded goods (automotive sight shields, fender extensions, toys, housewares), and adhesives (solvent cement and hot-melt types). The principal commercial supplier of styrenic block copolymers is Shell (Kraton). In addition, two joint ventures have been announced to produce these materials. Dow Chemical and Exxon have formed Dexco Polymers to produce systemic block copolymers for pressure-sensitive adhesive, asphalt (qv), and thermoplastic impact modifier markets. The second joint venture is EnArco, formed by ARCO Chemical Co. and Enimont. This venture will produce products for similar applications (84). The U.S. consumption of these materials as of 1989 was 139,000 t (85). These materials typically possess a high volume fraction of styrene and are priced between the low cost resins, polystyrene, polyethylene, etc and the high cost resins, cellulosic, clear ABS, polycarbonates, etc. These same materials are used as compatibilizers and impact modifiers in polymer blends. Here, they lower surface tension, decreasing domain size in two-phase blends, usually resulting in improved properties, such as impact strength and flexural modulus.

Polyurethane TPE exhibit toughness, low temperature flexibility, strength, and abrasion resistance. They are produced using a bulk or solution reaction of a polyol with a diisocyanate and are used largely in fabric coatings and injection molded and extruded goods (exterior automotive parts, gears, gaskets, etc). Important commercial producers include BASF (Elastollan), Bayer (Desmopan), Dow Chemical (Pellethane), and B. F. Goodrich (Estane). U.S. production is estimated at 21,000 t (86).

In contrast, the copolyester–ether block copolymer TPE are relatively expensive with high performance characteristics. These materials exhibit a two-phase morphology in which the hard crystalline segments made from polyester act as thermally reversible cross-links. The elastomeric character of the polymer arises from the amorphous soft polyether ester segments. They are produced by a melt-transesterification polymerization process and can be processed by conventional techniques such as injection, blow, transfer, or rotational molding. Important commercial products are produced by Du Pont (Hytrel), Eastman Kodak (Ecdel), General Electric (Lomod), and Hoechst Celanese (Riteflex). U.S. copolyester–ether consumption in 1988 was 12,000 t. Their most important uses are in wire cable materials (eg, the coiled stretch telephone cords), injection-molded articles (eg, small mechanical parts), and high pressure hoses (87).

Certain block copolymers have also found application as surfactants (88). For example, AB or ABA block copolymers in which one block is hydrophilic and one block is hydrophobic have proven useful for emulsifying aqueous and non-aqueous substances and for wetting the surface of materials. Examples of such surfactants are the poly(propylene oxide-*block*-ethylene oxide) materials, known as Pluronics (BASC Wyandotte Co.).

Graft Copolymers

Two commercially significant graft copolymers are acrylonitrile–butadiene–styrene (ABS) resins and impact polystyrene (IPS) plastics. Both of these families of materials were once simple mechanical polymer blends, but today such compositions are generally graft copolymers or blends of graft copolymers and homopolymers.

ABS is the sixth largest volume thermoplastic resin and the principal engineering (structural or load bearing) plastic (89). ABS is a terpolymer manufactured by copolymerizing acrylonitrile and styrene in the presence of polybutadiene rubber. Important producers of ABS plastics include General Electric, Monsanto (Lustran), and Dow (Abtec) (see ACRYLONITRILE POLYMERS).

The properties of ABS can be modified by varying the relative proportions of the basic components, the degree of grafting, and the molecular weight. For example, increasing the rubber content reduces tensile strength, modulus, and melt flow and increases impact strength. Rubber particle size also affects properties: small rubber particle size makes ABS stiffer and glossier but less tough. Larger particle size improves toughness but at the expense of stiffness and gloss. Combinations of good properties are usually found with particle sizes between 0.3 and 0.5 μm (90).

A variety of ABS grades are available with Izod impact strengths ranging from 106.7 J m (2 ft-lb in.) to 640.5 J m 12 ft-lb in., heat distortion temperatures from 80 to 115°C, and tensile strengths from 27.6 MPa (4000 psi) to 55.2 MPa (8000 psi). Special ABS resins with high flow characteristics for the injection molding of intricate parts with controlled gloss, or deep-draw vacuum formability can also be made. In addition, certain ABS plastics are available in pellet or powder form for use in blending with other polymers such as PVC. ABS plastics can be processed by all the techniques common to thermoplastics. The consumption of ABS in 1990 was 557,000 t (89). The principal markets for ABS are automotive (25% of total consumption), business machines (24%), and pipes and fittings (13%) (91).

Impact polystyrene (IPS) is one of a class of materials that contains rubber grafted with polystyrene. This composition is usually produced by polymerizing styrene (by mass or solution free-radical polymerization) in the presence of a small amount (ca 5%) of dissolved elastomer. Some of the important producers of impact-resistant polystyrenes are BASF (Polystyrol), Dow (Styron), and Monsanto (Lustrex). The 1988 U.S. production of impact polystyrene was more than 1 million t (92).

In these rubber-modified polystyrene polymers, the rubbers should have low *T_g*, large particle sizes (0.5–5μm), graftable and cross-linkable sites, and should be compatible with styrene monomer (93). Polybutadiene, with a *T_g* of -85°C, meets all of these requirements and is used most frequently. These rubber-modified systems exhibit excellent low temperature impact strength, a required attribute for use in refrigerators.

The small amount of dispersed and grafted rubber in the polystyrene glassy matrix results in tremendous improvements in IPS's resistance to brittle failure under high speed impact, without a sacrifice in mechanical or thermal properties. As the rubber content increases, impact strength increases, while rigidity, heat deflection temperature, and clarity gradually decrease. The exact mechanism for improving the impact strength of polystyrene with grafted rubbers is uncertain. A number of theories relating crack propagation, energy absorption, and shear band formation have been advanced (94). In addition, understanding the reasons for the ductility of the cross-linked rubber–PS blends must be addressed. Inter-particle distances and the mean free path

between the particles may be important in controlling fracture in IPS (95).
Impact polystyrenes have found use in the manufacture of refrigerator door liners and in packaging. They are also commonly used for appliance housings, radio and television cabinets, sporting goods, toys, cameras, furniture, luggage, pipe and fittings, automotive parts, and women's shoe heels (96).

Characterization of Copolymers

The characterization of copolymers must distinguish copolymers from polymer blends and the various types of copolymers from each other (97,98). In addition, the exact molecular structure, architecture, purity, supermolecular structure, and sequence distribution must be determined.
Assessing whether a material is a copolymer or a mixture of homopolymers can sometimes be accomplished by extracting the prospective copolymers with solvents selective for the component homopolymers. However, this method is effective only when the copolymer segments differ significantly in solubility behavior. The situation is further complicated by the fact that block copolymers can themselves act as compatibilizing agents for homopolymers and can confound extraction experiments. Alternatively, solution fractionation has been used to test copolymers versus homopolymers or polymer blends. Determining whether a co-polymer is block or random can often be made by ¹H nmr, ¹³C nmr, or from *T_g* measurements (differential scanning calorimetry (dsc) or thermal-mechanical tests). However, detailed monomer unit sequence distributions are usually best determined by ¹³C nmr. Using this technique, information on dyad (two monomer units) up to pentad (groups of five monomer units) can be obtained (99,100).

An indication of whether block or copolymer architecture is AB, ABA, or -(AB)-_n can often be seen in its rheological behavior. For example, ABA and -(AB)-_n block polymers exhibit higher melt viscosities than do AB diblock copolymers with similar molecular weights. The former two architectures partially preserve their physical network even in the melt. Gel permeation chromatography (gpc) is useful for determining the presence of homopolymer A and diblock AB impurities in anionically prepared ABA triblock copolymers.

The molecular structure of the copolymers is also important. Molecular-weight measurements (osmometry, gpc) and functional group analysis are useful. Block copolymers require supermolecular (morphological) structural information as well. A listing of typical copolymer characterization tools and methods is shown in Table 6.

Table 6. Copolymer Characterization ^a

Determination of	Method
copolymer vs homopolymer	solubility characteristics
	solid-phase extraction
	solution fractionation
	film clarity
	solution compatibility
	molecular weight distribution
	density-gradient ultracentrifugation
block vs random nature	gpc
	pmr
	ir
	dma
	dsc
	electron microscopy
	small-angle x-ray scattering
	mechanical properties
	rheological properties
	crystallinity
molecular structure	solution light scattering
	osmometry
	solution light scattering
	ultracentrifugation
	gpc
	solution viscometry
	oligomer analysis
	selective degradation
	elemental and functional group analysis
	elastic recovery
architecture and purity	rheological characteristics
	gpc
	density-gradient ultracentrifugation
supermolecular structure	dma
	dsc
	rheological characteristics
	electron microscopy and scanning electron microscopy
	wide-angle x-ray
	small-angle x-ray
	birefringence
	small-angle light scattering

^a Ref. 101.

Economic Aspects

The economic importance of copolymers can be clearly illustrated by a comparison of U.S. production of various homopolymer and copolymer elastomers and resins (102). Figure 5 shows the relative contribution of elastomeric copolymers (SBR, ethylene-propylene, nitrile rubber) and elastomeric homopolymers (polybutadiene, polyisoprene) to the total production of synthetic elastomers. Clearly, SBR, a random copolymer, constitutes the bulk of the entire U.S. production. Copolymers of ethylene and propylene, and nitrile rubber (a random copolymer of butadiene and acrylonitrile) are manufactured in smaller quantities. Nevertheless, the latter copolymers approach the volume of elastomeric butadiene homopolymers.

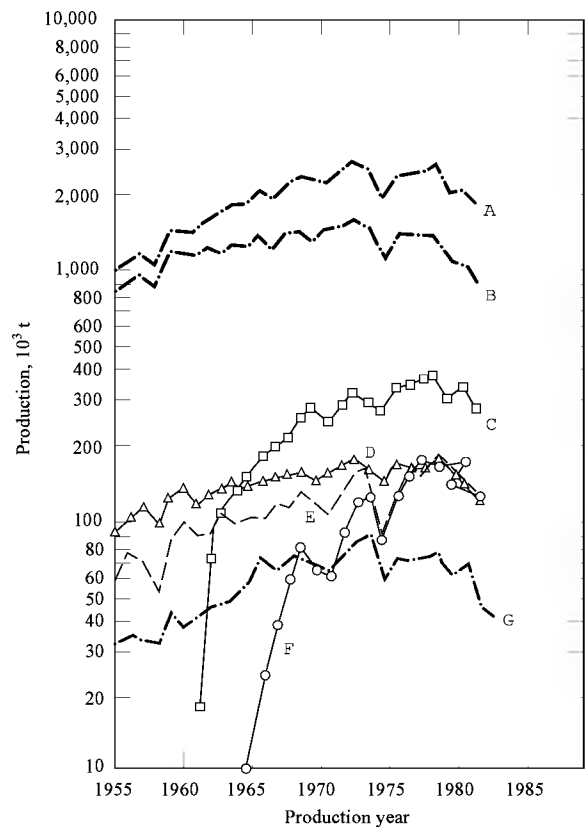


Fig. 5. U.S. production of synthetic elastomers (thousands of metric tons) versus production year (63). A, Total; B, SBR; C, polybutadiene; D, polychloroprene; E, butyl rubber; F, ethylene-propylene; G, nitrile.

The relative U.S. production of styrene homopolymer and copolymer resins is also noteworthy (103) (Fig. 6). The impact polystyrene (graft and polymer blend) copolymers are produced in nearly the same quantities as styrene homopolymers. The ABS resins are synthesized in lesser, yet significant, quantities.

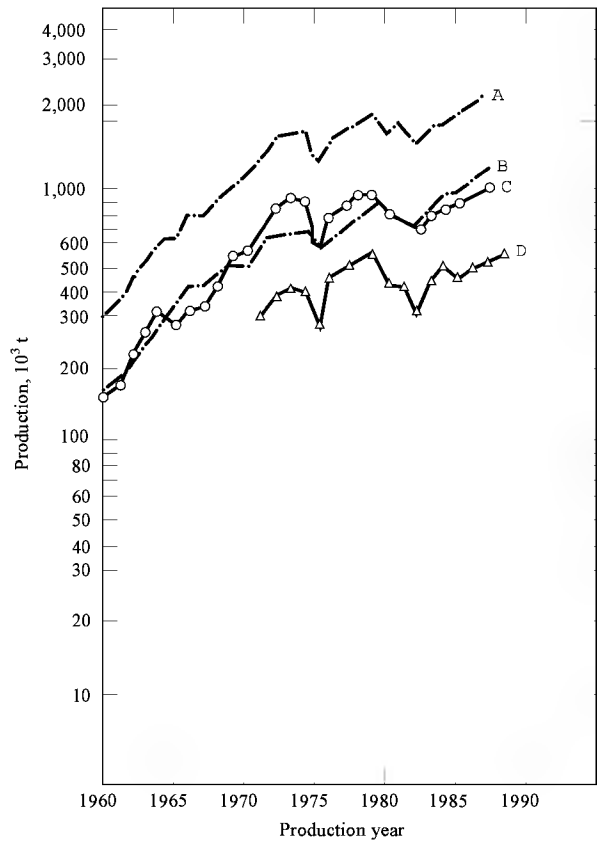


Fig. 6. U.S. production of polystyrene versus production year (103). A, Total; B, straight; C, impact (rubber modified); D, ABS.

Future Trends

Copolymer technology is progressing along two "fronts." First, new applications for copolymers are being found to increase the volume of materials that are already commercially available. One example of this is the rapid growth of styrenic block copolymers sold as asphalt (qv) and polymer modifiers over the past 10 years (Fig. 7). Another is the increased interest in graft and block copolymers as compatibilizers for polymer blends and alloys. Of particular interest are compatibilizers for recycled polymer scrap.

The second front originates in the polymer synthesis community. Efforts are mainly directed toward production of monodisperse block copolymers by living polymerizations. These structures typically result in microphase separated systems; if one block is a high *T_g* material and the other is elastomeric in

nature, then the overall system will be thermoplastic. Although living anionic polymerizations have been developed to synthesize copolymers (105), they are limited in the scope of the reaction with respect to polar monomers and tolerance of impurities, oxygen, and water. Du Pont has developed group transfer polymerization to make block copolymers of more polar monomers, such as acrylics (38,106). These polymerizations proceed at room temperature, yielding nearly monodisperse polymers, block copolymers, and difunctionalized materials (telechelics) (107). The telechelic materials can be used in block copolymers produced by step-growth polymerization. The first targeted applications for this technology are in the high solids coatings area (108).

Much academic work has centered on developing living catalyst systems that can produce monodisperse block copolymers by ring-opening olefin metathesis (ROMP) (109). Block copolymers of narrow molecular-weight distribution and telechelic starting materials for step-growth block copolymerization can be synthesized using this chemistry. Other transition-metal catalyzed living polymerization that form C–C bonds are known in addition to these ROMP catalysts. All of these systems are limited, however, to a small scope of monomers that can be employed (110–112). There has been a report of block copolymerization of butadiene and isocyanide monomers using $[\eta^3\text{-C}_3\text{H}_5\text{-Ni(OC(O)CF}_3)_2]$ (113). These block copolymers exhibit microphase separation with the helical polyisocyanide domains in the rubbery polybutadiene continuous phase.

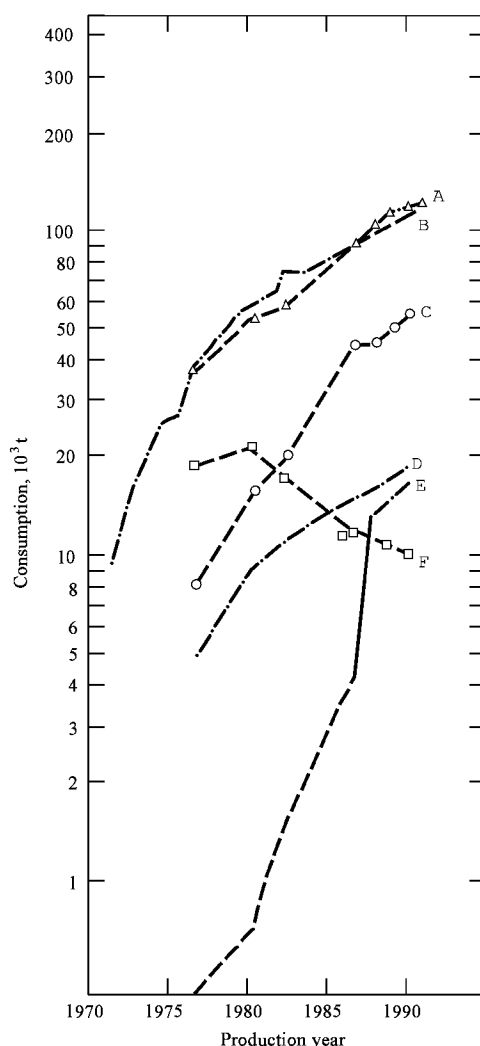


Fig. 7. U.S. production and consumption of styrene block copolymers (104). A, Total production; B, consumption; C, adhesives and sealants; D, polymer modifiers; E, asphalt modifiers; F, footwear.

Finally, copolymers of the future will be closely related to society's energy choices and monomer availability and price. Many of today's high volume copolymers owe their success to low cost petrochemical fuels and monomers. However, future shifts to coal, nuclear, or solar energy, as well as to renewable resources (eg, agriculture and paper pulp chemicals, etc) may cause new materials to capture the general-purpose market. High volume copolymers tend to be those generated by energy- (and cost-) efficient processes. The total energy requirement of the monomers, synthetic methods, and processing needs to be considered.

Furthermore, increased governmental scrutiny of chemical substances will make it more difficult to bring a new product to market. The choice of comonomers and copolymers may be based partly on EPA, FDA, OSHA, and TSCA rulings. In addition to these regulations, the thrust toward recycling polymers is expected to impact copolymer production. The ability to recover and reprocess these materials will be a key factor for economic success.

If new copolymers cannot be competitively introduced, new ways must be developed for improving old ones. This might involve copolymer blending or improved processing techniques. For example, the success of ABS–polycarbonate alloys (such as the Bayblend series of materials available from Miles) is an incentive for studying other copolymer blends (114). The importance of new processing methods for improved product economics has been verified by the introduction of reaction injection molding (RIM) and reactive extrusion. RIM is an efficient method for fabricating thermoplastic polyurethane copolymers (115). Reactive extrusion provides a means to make *in situ* grafted materials during the extrusion-processing step.

There will also be a growing need for macromolecules for special end use requirements. One method for obtaining these specialty materials is through copolymerization. Biomedical materials are of special interest. For example, biocompatible copolymers are becoming increasingly important. Applications of polyurethane copolymers (such as Ethicon's Biomer) to surgical devices, such as endotracheal tubes, synthetic blood vessels, heart valves, and even artificial hearts, suggest that specialty copolymers will play increasingly important roles in such applications (116).

BIBLIOGRAPHY

"Copolymers," in *ECT* 3rd ed., Vol. 6, pp. 798–818, by D. N. Schulz and D. P. Tate, Firestone Rubber & Tire Co.

1. G. Odian, *Principles of Polymerization*, 3rd ed., John Wiley & Sons, Inc., New York, 1991, p. 141.
2. IUPAC, *Pure Appl. Chem.* **57**, 1427 (1985).
3. N. M. Bikales in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, Vol. 10, John Wiley & Sons, Inc., New York, 1987, p. 191.
4. D. A. Tirrell, in Ref. 3, Vol. 4, pp. 192.

5. Ref. 1, pp. 453–523 and references therein.
6. F. R. Mayo and F. M. Lewis, *J. Am. Chem. Soc.* **66**, 1594 (1944).
7. T. Alfrey, Jr., and G. Goldfinger, *J. Chem. Phys.* **12**, 205 (1944).
8. T. Alfrey, Jr., and C. C. Price, *J. Polym. Sci.* **2**, 101 (1947).
9. M. Fineman and S. D. Ross, *J. Polym. Sci.* **5**, 259 (1950).
10. T. Kelen and F. Tudos, *J. Makromol. Sci. Chem.* **A9**, 1 (1975).
11. R. C. McFadane, P. M. Reilly, and K. F. O'Driscoll, *J. Polym. Sci. Polym. Chem. Ed.* **18**, 251 (1980).
12. G. C. Laurier, K. F. O'Driscoll, and P. M. Reilly, *J. Polym. Sci. Polym. Symp.* **72**, 171 (1985).
13. L. H. Garcia-Rubio, M. G. Lord, J. F. MacGregor, and A. E. Hamielic, *Polymer* **26**, 2001 (1985).
14. H. Patino-Leal, P. M. Reilly, and K. F. O'Driscoll, *J. Polym. Lett. Ed.* **18**, 219 (1980).
15. Ref. 1, p. 431.
16. R. W. Lenz, *Organic Chemistry of High Polymers*, Wiley-Interscience, New York, 1967, p. 383.
17. I. Crescentini, G. B. Getchele, and A. Zanella, *J. Appl. Polym. Sci.* **9**, 1323 (1965).
18. T. Nakata, T. Otsu, and M. Imoto, *J. Polym. Sci.* **A3**, 3383 (1965).
19. T. Alfrey, Jr., and E. Lavin, *J. Am. Chem. Soc.* **67**, 2044 (1945).
20. T. Alfrey, Jr., J. Bohrer, H. Haas, and C. Lewis, *J. Polym. Sci.* **5**, 719 (1950).
21. P. Argon, T. Alfrey, Jr., J. Bohrer, H. Haas, and H. Wechsler, *J. Polym. Sci.* **3**, 159 (1948).
22. G. Mazzanti, A. Valvassori, and G. Pajaro, *Chem. Ind. (Milan)* **39**, 743 (1957).
23. M. Ratzsch, W. Schneider, and D. Musche, *J. Polym. Sci. Part A* **9**, 785 (1971).
24. L. Yu and co-workers, *J. Polym. Sci.* **58**, 1181 (1962).
25. A. A. Korotkov, S. P. Mitzeneggndler, and K. M. Aleyev, *Polym. Sci. U.S.S.R.* **3**, 487 (1962).
26. N. Ashikari, *Bull. Chem. Soc. Jpn.* **32**, 1060 (1959).
27. W. Saltman, in M. Morton, ed., *Rubber Technology*, 2nd ed., Van Nostrand Reinhold Co., Inc., New York, 1973, Chapt. 7.
28. W. G. Barb, *J. Am. Chem. Soc.* **75**, 224 (1953).
29. Ref. 1, p. 476.
30. L. E. Forman, in J. P. Kennedy and E. G. Tornquist, eds., *Polymer Chemistry of Synthetic Elastomers II, High Polymers*, Vol. 23, Wiley-Interscience, New York, 1969.
31. L. J. Fetters, *J. Elastoplast.* **4**, 34 (1972).
32. A. F. Halasa in A. W. Langer, ed., *Polyamine-Chelated Alkali Metal Compounds, ACS Advances in Chemistry*, No. **130**, American Chemical Society, Washington, D.C., 1974.
33. J. M. G. Cowie, in G. A. Allen and J. C. Bevington, eds., *Comprehensive Polymer Science: The Synthesis, Characterization, Reactions, and Applications of Polymers*, Vol. 3, Pergamon Press, New York, 1989, p. 33.
34. R. L. Zapp and P. Haus, in Ref. 27, Chapt. 10.
35. P. Dreyfus and J. P. Kennedy, *J. Polym. Sci. Part C* **56**, 129 (1976).
36. T. G. Heggs, in D. C. Allport and W. H. Jones, eds., *Block Copolymers*, John Wiley & Sons, Inc., New York, 1973, Chapt. 3.
37. D. C. Allport and A. A. Mohajer, in Ref. 33, Chapt. 5.
38. D. Y. Sogah, W. R. Hertler, O. W. Webster, and G. M. Cohen, *Macromolecules* **20**, 1473 (1987).
39. A. Noshay and J. E. McGrath, *Block Copolymers—Overview and Critical Survey*, Academic Press, Inc. New York, 1977, pp. 139–140.
40. S. Inoue, *Chemtech.* **6**, 588 (1976).
41. L. F. Cannizzo and R. H. Grubbs, *Macromolecules* **21**, 1961 (1988).
42. R. Milkovich, in J. E. McGrath, ed., *Anionic Polymerization Kinetics, Mechanisms, and Synthesis, ACS Symposium Series*, No. **166**, American Chemical Society, Washington, D.C., 1981, p. 41.
43. P. Rempp, E. Franta, and J. E. Herz, in *Advances in Polymer Science* **86**, Springer-Verlag, New York, 1988, pp. 145–173.
44. F. Halverson, J. E. Lancaster, and M. N. O'Conner, *Macromolecules* **18**, 1139 (1985).
45. S. H. Rose, *J. Polym. Sci. Part B* **6**, 837 (1968).
46. G. S. Kyker and T. A. Antkowiak, *Rubber Chem. Technol.* **47**, 32 (1974).
47. O. B. Edgar and R. Hill, *J. Polym. Sci.* **8**, 1 (1952).
48. *Firestone Product Specialized Elastomers, M-010-65*, Firestone Synthetic Rubber and Latex Co., Akron, Ohio, 1973.
49. T. G. Fox, *Bull. Am. Phys. Soc.* **1**, 123 (1956).
50. S. R. Sandler and W. Karo, *Polymer Synthesis*, Vol. 1, Academic Press, Inc., New York, 1974, p. 520.
51. F. W. Billmeyer, Jr., *Textbook of Polymer Science*, 3rd ed., Wiley-Interscience, New York, 1984, p. 348.
52. C. F. Hammer, *Macromolecules* **4**, 69 (1971).
53. G. ten Brinke, F. E. Karasz, and W. J. MacKnight, *Macromolecules* **16**, 1827 (1983).
54. D. Rigby, J. L. Lin, and R. J. Roe, *Macromolecules* **18**, 2269 (1985).
55. L. A. Utracki, *Polymer Alloys and Blends, Thermodynamics and Rheology*, Hanser Publishers, New York, 1989, p. 124, and references cited therein.
56. G. E. Molau, in S. L. Agarawal, ed., *Block Copolymers*, Plenum Press, New York, 1970, p. 79.
57. M. Matsuo, T. Ueno, H. Horino, S. Chuiyo, and H. Asai, *Polymer* **9**, 425 (1968).
58. H. C. Kan, J. D. Ferry, and L. J. Fetters, *Macromolecules* **13**, 1571 (1980).
59. R. Fayt, R. Jerome, and J. Teyssie, *J. Polym. Sci. Part B* **27**(4), 775 (1989).
60. L. Liebler, *Makromol. Chem. Macromol. Symp.* **16**, 1 (1988).
61. P. M. Subramanian in *TAPPI Polymers, Laminations & Coatings Conf. Proc.*, **1**, TAPPI, Atlanta, Ga., 1984, p. 341.
62. C. A. Harper, in C. A. Harper, ed., *Handbook of Plastics and Elastomers*, McGraw-Hill Book Co., New York, 1975, Chapt. 1.
63. J. R. Hardy, *Chemical Economics Handbook, 1990 Directory of Chemical Producers (United States)*, Stanford Research Institute, Menlo Park, Calif., p. 913.
64. J. P. Morrill, in Ref. 27, Chapt. 12.
65. Ref. 63, p. 598.
66. K. Hashimoto and Y. Todani, in A. K. Bhowick and H. L. Stephens, eds., *Handbook of Elastomers—New Developments and Technology*, Marcel Dekker, Inc., New York, 1988, p. 741.
67. E. L. Borg, in Ref. 27, Chapt. 9.
68. Ref. 63, p. 597.
69. D. Caldwell, in R. Juran, ed., *Modern Plastics Encyclopedia*, Vol. 11, Issue 67, McGraw-Hill, Book Co., Inc., New York, 1991, p. 101.
70. M. J. Howard, ed., *The International Plastics Selector, 1977*, International Plastics Selector, Inc. of Codura Publications, La Jolla, Calif., 1977, p. 544.
71. Ref. 63, p. 912.
72. H. E. Frey, *CEH Marketing Research Report, Polyvinyl Chloride Resins*, Stanford Research Institute, Menlo Park, Calif., 1988, p. 580.1881G.
73. Ref. 72, p. 580.1881O.
74. Ref. 63, p. 894.
75. . H. G. Elias and F. Vohwinkel, *New Commercial Polymers 2*, Gordon & Breach Publishers, New York, 1986, p. 160.
76. J. F. Auchter, R. Mulach, and K. Tsuchiya, *CEH Data Summary, Fluoroelastomers*, Stanford Research Institute, Menlo Park, Calif., 1989, p.

525.6000A.

77. J. Furukawa, in J. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, 2nd ed., Vol. 4, John Wiley & Sons, New York, 1986, p. 233.

78. **U.S. Pat. 3,737,417** (1973), K. Hayashi, A. Kawaski, and I. Maruyama.

79. A. Kawasaki and co-workers, *International Rubber Conference, Tokyo*, 1975, p. 13.

80. A. Kawasaki and co-workers, *J. Soc. Rubber Ind. Jpn.*, **49**, 169, (1976).

81. C. P. Rader, in Ref. 39, pp. 69–75.

82. H. E. Frey and A. J. Wolfe, in *Chemical Economics Handbook*, Stanford Research Institute, Menlo Park, Calif., 1977, p. 580.1502D.

83. Ref. 39, p. 71.

84. L. Brodrecht, P. Mortensen, and M. Tachiro, *Styrenic Thermoplastic Elastomers*, Stanford Research Institute, Menlo Park, Calif., 1990, p. 525.8600A.

85. Ref. 84, p. 525.8600G.

86. E. Linak, F. P. Kalt, and Y. Sakuma, *Polyurethane Elastomers, CEH Marketing Research Report*, Stanford Research Institute, Menlo Park, Calif., July 1990, p. 525.6600A.

87. L. Brodrecht, R. Mulach, and K. Tsuchiya, *Copolyester-ether Elastomers, CEH Data Summary*, Stanford Research Institute, Menlo Park, Calif., June 1989, p. 525.8200A.

88. Ref. 39, pp. 69–75.

89. *Mod. Plastics*, **68**, 113 (Jan. 1991).

90. A. Echte, in C. K. Riew, ed., *Rubber-Toughened Plastics, Advances in Chemistry Series*, No. 222, American Chemical Society, Washington, D.C., 1989.

91. H. E. Frey and S. Stai, *CEH Marketing Research Report, Acrylonitrile–Butadiene–Styrene Resins*, Stanford Research Institute, Menlo Park, Calif., p. 580.0180E.

92. H. E. Frey and S. Stai, *CEH Marketing Research Report, Polystyrene*, Stanford Research Institute, Menlo Park, Calif., 1988, p. 580.1501N.

93. N. Platzer, *Chemtech.* **7**(10), 634 (1977).

94. R. D. Deanin, A. A. Deanin, and T. Sjoblom in L. H. Sperling, ed., *Recent Advances in Polymer Blends, Grafts, and Blocks*, Plenum Press, New York, 1974, Chapt. 2.

95. G. H. Michler, *Plaste Kautschuk* **26**, 680 (1986).

96. Ref. 92, p. 580.1501Q.

97. Ref. 39, p. 48.

98. D. Braun, H. Cherdon, and W. Kern, *Techniques of Polymer Synthesis and Characterization*, Wiley-Interscience, New York, 1971, p. 88.

99. J. C. Randall, *Polymer Sequence Determination—Carbon 13 NMR Methods*, Academic Press, New York, 1977.

100. J. C. Randall, ed., *NMR and Macromolecules, Sequence, Dynamic, and Domain Structure, ACS Symposium Series*, No. **247**, American Chemical Society, Washington, D.C., 1984.

101. Ref. 39, p. 50.

102. I. J. Freeman, *Elastomers, U.S. Data Summary, Chemical Economics Handbook*, Stanford Research Institute, Menlo Park, Calif., 1984, p. 525.1000D.

103. Ref. 92, p. 580.1501M.

104. Ref. 84, p. 525.8600I.

105. R. P. Quirk, D. J. Kinning, and L. J. Fetters, in G. Allen, ed., *Block Copolymers, in Comprehensive Polymer Science: The Synthesis, Characterization, Reactions & Applications of Polymers. Vol. 7, Specialty Polymers and Polymer Processing*, Pergamon Press, New York, 1989.

106. O. W. Webster, W. R. Hertler, D. Y. Sogah, W. B. Farnham, and T. V. Rajan Babu, *J. Am. Chem. Soc.* **105**, 5706 (1983).

107. D. Y. Sogah and O. W. Webster, *J. Polym. Sci. Polym. Lett.* **21**, 927 (1983).

108. D. B. Parkinson, *Group-Transfer Polymerization Applications, PEP Review No. 83-2-3*, Stanford Research Institute, Menlo Park, Calif., 1984, p. 9.

109. R. Grubbs and B. M. Novak, in J. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, Supplement, John Wiley & Sons, Inc., New York, 1985, p. 420.

110. T. Yoshimura, T. Masuda, and T. Higashimura, *Macromolecules* **21**, 1899 (1988).

111. Y. Doi, S. Suzuki, and K. Soga, *Macromolecules* **19**, 2896 (1986).

112. P. Hadjiandreou, M. Julemont, and P. Tessyie, *Macromolecules* **17**, 2455 (1984).

113. T. J. Deming and B. M. Novak, *Macromolecules* **24**, 5478 (1991).

114. S. Y. Kienzle, *Polymer Blends and Alloys, Guidebook to Commercial Products*, Technomic Publishing Co., Lancaster, Pa., 1988, p. 363.

115. *Rubber Plastic News* **6**(26), 16 (July 25, 1977).

116. M. D. Lelah and S. L. Cooper, *Polyurethanes in Medicine*, CRC Press, Boca Raton, Fla., 1986, p. 57.

Christine A. Costello
Donald N. Schulz
Exxon Research and Engineering Company

COPPER

Copper [7440-50-8], Cu, critically important to the development of civilization, is the only common metal found naturally in the metallic state. It was thus suitable for the production of tools, and ancient people made use of its easy workability and beauty. Furthermore, the ease with which the oxide can be reduced to the metal, together with the tendency of copper to alloy with other metals naturally present in the ores, promoted broad usage.

Copper, the first element of Group 11 (IB) of the Periodic Table, is immediately above silver and gold. It is classed with silver and gold as a noble metal and can be found in nature in the elemental form. Copper occurs as two natural isotopes, ³Cu and ⁶⁵Cu (1).

The high electrical and thermal conductivities and corrosion resistance of copper combined with its workability give the metal its very wide range of commercial applications. Unlike most metals, which are alloyed with other elements to enhance properties, for example, alloy steel and aluminum, copper is primarily used in its pure, unalloyed form.

The earliest recorded use of copper was in northern Iraq about 8500 BC and then in Asia Minor and Egypt around 7000 BC. Copper items found on the Sinai Peninsula have been dated at about 3800 BC (2). The world's oldest known copper smelting furnace, dating to 3500 BC, is located near the Timna copper mine in Israel. Deposits on Cyprus were mined as early as 3000 BC. During this early period the Egyptians developed the metallurgical arts and the use of bronze became relatively common. The mines on Cyprus were prized possessions of the empires that followed the Egyptians and became the chief source of metal for the Roman Empire. The metal was named *aes syprium* and subsequently *cuprum*, from which is derived the English word *copper* and the symbol Cu.

Developments in the metallurgy of copper or its alloys were mentioned in 1556 in *De Re Metallica*, where the process of copper ore was described by Agricola (see also COPPER ALLOYS). About that time, smelting operations commenced at Mansfield, Germany, and at the Swansea smelter in Wales. Both smelters employed successive oxidations and reductions to eliminate iron and sulfur. The process used in the Swansea smelter is similar to modern techniques.

Between 1869 and 1877 the Calumet and Hecla Company in Michigan became the largest individual copper producer in the world; its annual copper production was less than 6200 t. By 1877, the mines of Rio Tinto in the Huelva province of southern Spain took a leading position, with an annual output of slightly more than 24,000 t. In the 1890s the Anaconda Copper Company, Montana, surpassed this mark with an annual output of 34,000 t. The Anaconda mine maintained its position as the world's largest copper mine until 1920, increasing annual output to more than 50,000 t.

Large-scale mining of low grade ores is a development of the twentieth century. The potential of massive low grade porphyry deposits was first realized when concentration methods were developed at the open-pit mine at Bingham Canyon, Utah (3). Introduction of froth flotation (qv) for beneficiation of sulfide ores during the 1920s improved metal recovery and gave impetus to the exploitation of the low grade porphyry deposits in Arizona, which became the leading copper producing area in the United States.

The exploitation of large ore bodies in Chile and Peru has made South America the world's largest producer of copper. The United States is the second largest, followed by Zaire and Zambia, and the CIS. Other important deposits are found in southern Oceania (Papua New Guinea, the Philippines, and Indonesia), Canada, Mexico, and Poland.

Occurrence

Cosmically, copper is relatively abundant: 100–400 ppm are found in the metal phase of meteoric iron (see EXTRATERRESTRIAL MATERIALS). The high affinity of copper for sulfur is the principal factor in determining the manner of occurrence in the earth's crust. Copper–iron sulfides are the last minerals to crystallize, and these fill the interstices between other minerals in igneous rocks, which contain an average of about 60–70 ppm copper (4). Other copper compounds occurring in nature are oxides and silicates. The strong affinity of copper for sulfur is the prime factor in separating copper from iron in the pyrometallurgical reduction of copper from sulfide ore.

Copper ore minerals may be classified as primary, secondary, oxidized, and native copper. Primary minerals were concentrated in ore bodies by hydrothermal processes; secondary minerals formed when copper sulfide deposits exposed at the surface were leached by weathering and groundwater, and the copper reprecipitated near the water table (see METALLURGY, EXTRACTIVE). The important copper minerals are listed in Table 1. Of the sulfide ores, bornite, chalcopyrite, and tetrahedrite–tennantite are primary minerals and covellite, chalcocite, and digenite are more commonly secondary minerals. The oxide minerals, such as chrysocolla, malachite, and azurite, were formed by oxidation of surface sulfides. Native copper is usually found in the oxidized zone. However, the principal native copper deposits in Michigan are considered primary (5).

Table 1. Important Copper Minerals^a

Mineral	CAS Registry Number	Composition	Copper, %	Color	Crystal system	Luster	Mohs' hardness	Specific gravity
<i>Sulfides</i>								
bornite	[1308-82-3]	Cu ₅ FeS ₄	63.3	between copper red and brown	isometric	metallic	3	5.06–5.08
chalcopyrite	[1308-56-7]	CuFeS ₂	34.5	brass yellow	tetragonal	metallic	3.5–4	4.1–4.3
tetrahedrite ^b	[1317-91-5]	Cu ₁₂ Sb ₄ S ₁₃	45.8	flint gray to iron black	isometric	metallic (splendent)	3–4.5	4.6
tennantite ^b	[12178-49-3]	Cu ₁₂ As ₄ S ₁₃	51.6	blackish lead gray to iron black	isometric	metallic	3–4.5	4.37–4.49
chalcocite	[21112-20-4]	Cu ₂ S	79.8	blackish lead gray	orthorhombic	metallic	2.5–3	5.5–5.8
covellite	[19138-68-2]	CuS	66.4	indigo blue or darker	hexagonal	submetallic to resinous	1.5–2	4.6–4.76
<i>Oxides</i>								
cuprite	[1308-76-5]	Cu ₂ O	88.8	red	isometric	adamantine to earthy	3.5–4	6.14
tenorite	[1317-92-6]	CuO	79.9	gray to black	monoclinic	metallic	3.5	5.8–4.03
malachite	[1319-53-5]	CuCO ₃ ·Cu(OH) ₂	57.3	bright green	monoclinic	adamantine to earthy	3.5–4	3.9–4.03
azurite	[1319-45-5]	2CuCO ₃ ·Cu(OH) ₃	55.1	azure blue	monoclinic	vitreous, almost adamantine	3.5–4	3.77–3.89
brochantite	[12068-81-4]	Cu ₄ SO ₄ (OH)	56.2	green	monoclinic	vitreous	3.5–4	3.9
atacamite	[1306-85-0]	Cu ₂ Cl(OH) ₃	59.5	green	orthorhombic	adamantine to vitreous	3–3.5	3.76–3.78

chrysocolla	[26318-99-0]	CuSiO ₃ ·2H ₂ O	36.0	green to blue	orthorhombic	vitreous to earthy	2.4	2.0–2.4
native copper	[7440-50-8]	Cu	100.0	copper red	isometric	metallic	2.5–3	8.95

^a Refs. 5–6.
^b These are the limits of a series having all possible intergrades.

Most copper deposits are (1) porphyry deposits and vein replacement deposits, (2) strata-bound deposits in sedimentary rocks, (3) massive sulfide deposits in volcanic rocks, (4) magmatic segregates associated with nickel in mafic intrusives, or (5) native copper, typified by the lava-associated deposits of the Keweenaw Peninsula, Michigan.

Almost two-thirds of the world's copper resources are porphyry deposits. The term porphyry is generally applied to a type of disseminated copper deposit that is hydrothermal in origin and characterized by a large proportion of minerals uniformly distributed as disseminations or in fractures and small veins. Copper contents are generally 1% or less. The most extensive porphyry deposits are located in western Canada, the southwestern United States, Mexico, and western South America. In addition to the porphyrys, there are large bedded copper deposits in Germany, Poland, the CIS, Australia, and central Africa.

In spite of low copper contents, massive horizontal development renders porphyry deposits amenable to large-scale production methods. Porphyry deposits are associated with igneous activity and intrusion of molten rocks into cooler parts of the earth's crust, often in connection with the formation of mountains. Erosion of mountainous areas exposes these deposits to weathering, and, under the right conditions, enables the formation of oxidized or secondary copper deposits. Copper mines in the United States are listed in Table 2.

Table 2. Leading Copper-Producing Mines in the United States in 1989^a

Ran k	Mine	County and State	Operator	Source of copper	Capacity, t × 10 ³
1	Morenci	Greenlee, Ariz.	Phelps Dodge Corp.	} copper–molybdenum ore, concentrated and leached	305
2	Bingham Canyon	Salt Lake, Utah	Kennecott, Utah Copper Corp.		245
3	Tyrone	Grant, N.M.	Phelps Dodge Corp. and Burro Chief Copper Co.	copper ore, concentrated and leached	160
4	Chino	Grant, N.M.	Phelps Dodge Corp.	} copper–molybdenum ore, concentrated and leached	150
5	San Manuel	Pinal, Ariz.	Magma Copper Co.		130
6	Ray	Pinal, Ariz.	ASARCO Incorporated	} copper ore, concentrated and leached	125
7	Sierrita	Pima, Ariz.	Cyprus Sierrita Corp.		117
8	Bagdad	Yavapai, Ariz.	Cyprus Bagdad Copper Co.	} copper–molybdenum ore, concentrated and leached	97
9	Pinto Valley	Gila, Ariz.	Pinto Valley Copper Corp.		92
10	Inspiration	Gila, Ariz.	Cyprus Miami Mining Corp.	} copper ore, leached	65
11	Mission Complex	Pima, Ariz.	ASARCO Incorporated		70
12	White Pine	Ontonagon, Mich.	Copper Range Co.	} copper ore, concentrated	50
13	Continental	Silver Bow, Mont.	Montana Resources Inc.	copper–molybdenum ore, concentrated	50
14	Troy	Lincoln, Mont.	ASARCO Incorporated	copper–silver ore, concentrated	18
15	San Xavier	Pima, Ariz.	ASARCO Incorporated	copper ore, concentrated	15
16	Casteel	Iron, Mo.	The Doe Run Co.	lead–copper ore, concentrated	20
17	Twin Buttes	Pima, Ariz.	Cyprus Sierrita Corp.		
18	Miami	Gila, Ariz.	Pinto Valley Copper Corp.	} copper ore, leached	6
19	Silver Bell	Pima, Ariz.	ASARCO Incorporated		5
20	Pinos Altos	Grant, N.M.	Cyprus Pinos Altos Corp.	copper ore, concentrated	8
21	Buick	Iron, Mo.	The Doe Run Co.	lead–zinc ore, concentrated	
22	Lakeshore	Pinal, Ariz.	Cyprus Casa Grande Corp.	copper ore, leached	5
23	Magmont	Iron, Mo.	Cominco American Incorporated	lead–zinc ore, concentrated	
24	Viburnum Vo. 28	Iron, Mo.	The Doe Run Co.	lead–copper ore, concentrated	
25	Fletcher	Reynolds, Mo.	The Doe Run Co.	lead–zinc ore, concentrated	

^a Ref. 7.

Exploration. Because it takes years to bring a mine into production, significant new copper deposits are sought and known reserves are expanded more or less continually. These exploration expenditures are highly sensitive to metal market conditions, and, as of this writing, gold is at a better price level than copper. Worddwide exploration continues, however, even after discovery of a deposit and the start of mining.

Ocean Nodules. A less conventional copper resource consists of deep-sea ferromanganese nodules. These nodules are primarily manganese, but some deposits contain over 1% copper. The nodules occur at many ocean sites, but the most valuable deposits are found in the Pacific Ocean. Although a number of companies are studying methods for recovering values from this source, copper resources from nodules must be considered tentative. World resources are estimated at 0.7 billion metric tons (8) (see OCEAN RAW MATERIALS).

A summary of the world's estimated primary copper resources is presented in Table 3.

Table 3. World Copper Resources, t × 10⁶ Copper^a

Location	Reserves	Reserve base ^b
North America		
United States	57	90

Canada	17	23
Mexico	17	23
other	1	15
<i>Total</i>	<i>90</i>	<i>150</i>
South America		
Chile	79	97
Peru	12	32
other	3	12
<i>Total</i>	<i>95</i>	<i>140</i>
Europe	50	70
Africa		
Zaire	26	30
Zambia	30	34
other	4	7
<i>Total</i>	<i>60</i>	<i>70</i>
Asia		
Philippines	12	18
other	14	19
<i>Total</i>	<i>25</i>	<i>35</i>
Oceania		
Australia	8	16
Papua, New Guinea	6	14
other	1	4
<i>Total</i>	<i>15</i>	<i>35</i>
<i>World total</i>	<i>340</i>	<i>500</i>

^a Ref. 9.

^b Includes reserves.

^c Totals may be rounded.

Properties

The properties of copper, silver, and gold are compared in Table 4. Like silver and gold, copper has an atomic structure that results in outstanding electrical and thermal conductivities and a high degree of malleability (see GOLD AND GOLD COMPOUNDS; SILVER AND SILVER ALLOYS). Although the ground state electronic configuration, $1s^22s^2p\ 3s^23p\ 3d^{10}4s^1$, implies a stable closed shell of 18 electrons, the shell is not inert. Rather, the underlying d orbitals appear to participate in metallic bonding by promotion of at least one d electron into a higher energy orbital of the outermost principal quantum level. There this electron is available for participation in electrical and thermal conduction (11). Although commercial copper is of excellent purity, differences in data obtained for pure copper and for electrolytic copper are significant.

Table 4. Properties of Pure Copper, Silver, and Gold^{a, b}

Property	Copper	Silver	Gold
CAS Registry Number	[7440-50-8]	[7440-22-4]	[7440-57-5]
atomic weight	63.54	107.87	196.97
atomic volume, cm ³ mol	7.11	10.27	10.22
mass numbers stable	63 (69.1)	107 (51.35)	197 (100)
isotopes (relative abundance, %)	65 (30.9)	109 (48.65)	
oxidation states	1, 2, 3	1, 2, 3	1, 2, 3
standard electrode potential, ^c	Cu Cu ⁺ 0.520	Ag Ag ⁺ 0.799	Au Au ⁺ 1.692
25°C, V	Cu Cu ²⁺ 0.337		
density, kg m ³	8.96 × 10 ³	10.49 × 10 ³	19.32 × 10 ³
metallic radius, nm ^d	0.1276	0.1442	0.1439
ionic radius, M ⁺ , nm	0.096	0.126	0.137
covalent radius, nm ^d	0.138	0.153	0.150
electronegativity ^d	2.43	2.30	2.88
ionization energy ^d , kJ mol ^e			
1st	745	732	891
2nd	1950	2070	
heat of atomization ^d , kJ mol ^e	339	286	354
thermal conductivity, W (m·K)	394	427	289
electrical resistivity at 20°C, μΩ cm	1.6730	1.59	2.35
temperature coeff. of electrical resistivity ^f	0.0068	0.0041	0.004
melting point, °C	1083	960.8	1063
heat of fusion, kJ kg ^e	212	102	67.4
boiling point, °C	2595	2212	2970
heat of vaporization, kJ kg ^e	7369	2400	1860
specific heat at 20°C, J (kg·°C) ^e	384	233	131 (18°C)
linear coeff. of expansion × 10 ⁶ per °C at 20°C	16.5	10.68	14.2
tensile strength, annealed metal, MPa ^g	230	280	170
modulus of elasticity, for hard drawn metal, MPa ^g	10.2–12 × 10 ⁴	7.75 × 10 ⁴	7.85 × 10 ⁴
modulus of rigidity, MPa ^g	44,000		
magnetic susceptibility at 18°C, cgs units g	0.086 × 10 ⁻⁶	0.20 × 10 ⁻⁶	0.15 × 10 ⁻⁶
emissivity, unoxidized metal at 100°C	0.03	0.052	0.02 ^h
viscosity at 1145°C, mPa·s(cP) ⁱ	3.41		

surface tension at 1150°C, mN · m (· dyn · cm) ⁱ	1104
^a Ref. 10.	
^b The crystal structure of each of the pure metals is fcc.	
^c Ref. 3.	
^d Ref. 11.	
^e To convert J to cal, divide by 4.184.	
^f From 0–100°C.	
^g To convert MPa to psi, multiply by 145.	
^h This value is not known with certainty.	
ⁱ Ref. 12.	

The unique nature of the electronic configuration of copper, which contributes to its high electrical and heat conductivity, also provides chemical properties intermediate between transition and 18-shell elements. Copper can give up the 4*s* electron to form the copper(I) ion [17493-86-6] or release an additional electron from the 3*d* orbitals to form the copper(II) ion [15158-11-9].

The higher ionization energy and smaller ionic radius of copper contribute to its forming oxides much less polar, less stable, and less basic than those of the alkali metals (13). Because of the relative instability of its oxides, copper joins silver in occurring in nature in the metallic state.

The standard oxidation potentials of copper(I) and copper(II) at 25°C are (14):



Copper(I) forms compounds with the anions of both strong and weak acids. Many of these compounds are stable and insoluble in water. Compounds and complexes of copper(I) are almost colorless because the inner 3*d* orbital of the copper is completely filled. There is a very strong tendency for copper(I) to disproportionate in aqueous solutions into copper(II) and metallic copper.



The properties of copper(II) are quite different. Ligands that form strong coordinate bonds bind copper(II) readily to form complexes in which the copper has coordination numbers of 4 or 6, such as tetraammine copper(II) [16828-95-8], [Cu(NH₃)₄]²⁺, and hexaaquacopper(II) [14946-74-8], [Cu(H₂O)₆]²⁺ (see COORDINATION COMPOUNDS). Formation of copper(II) complexes in aqueous solution depends on the ability of the ligands to compete with water for coordination sites. Most copper(II) complexes are colored and paramagnetic as a result of the unpaired electron in the 3*d* orbital (see COPPER COMPOUNDS).

Elemental copper is resistant to aerated alkaline solutions except in the presence of ammonia. Copper does not displace hydrogen from acid but dissolves readily in oxidizing acids such as nitric acid (qv) or in acid solutions that contain an oxidizing agent, such as sulfuric acid solution containing ferric sulfate. Because of corrosion resistance to salt solutions, copper is used in marine applications (see COATINGS, MARINE; CORROSION AND CORROSION CONTROL). Resistance to oxidation by water vapor at high temperatures has made copper a material of choice in cooling systems. Although the surface of copper oxidizes on exposure to the atmosphere, further corrosion is inhibited by the formation of a tightly adherent protective coating of corrosion products. In many instances this takes the form of a green patina that imparts a rich appearance to architectural and artistic uses. Studies of the chemistry of atmospheric corrosion have shown that the initial attack on copper metal involves the formation of sulfides and oxides. Upon further oxidation and reaction with water, basic copper sulfates, such as CuSO₄ · Cu(OH)₂ and CuSO₄ · 3Cu(OH)₂, are formed. The latter compound is found in nature as the mineral brochantite. A basic carbonate, CuCO₃ · Cu(OH)₂, may also be present in some deposits (15).

Recovery and Processing

Most copper is processed using a combination of mining, concentrating, smelting, and refining, or by leaching waste and solvent extraction–electrowinning. Open-pit mining is more common than underground mining, and the overburden, or waste, contains some copper. Frequently the waste is leached to extract the copper, which may be recovered by contacting the solution with an organic solvent (solvent extractor) and then recovering the copper by electroplating (qv) it on copper sheets (electrowinning).

The copper ore from the mine, often containing less than 1% copper, is transported to the concentrator, where it is crushed and then ground with water. The ground-ore slurry enters flotation cells, where copper minerals are collected as a froth known as concentrate (see FLOTATION). Following dewatering (qv) by filtration (qv), the minerals are shipped to the smelter. In the smelter the sulfide minerals react with oxygen and fluxes to produce impure copper metal, sulfur dioxide [7446-09-5], SO₂, and slag. Smelting occurs in two steps. In the smelting furnace, the copper concentrate reacts with air and oxygen, it is melted to produce matte, the mixed sulfides of copper and iron, sulfur dioxide gas, and a slag containing much of the iron and all of the silicates in the concentrates. In the next step, converting air is blown through the matte, producing impure copper, more sulfur dioxide gas, plus a slag containing the remaining iron. The impure copper is further fire-refined to adjust the oxygen and sulfur content, then cast into anodes, and purified by plating onto pure copper in an electrolytic tank house. Figure 1 illustrates these steps.

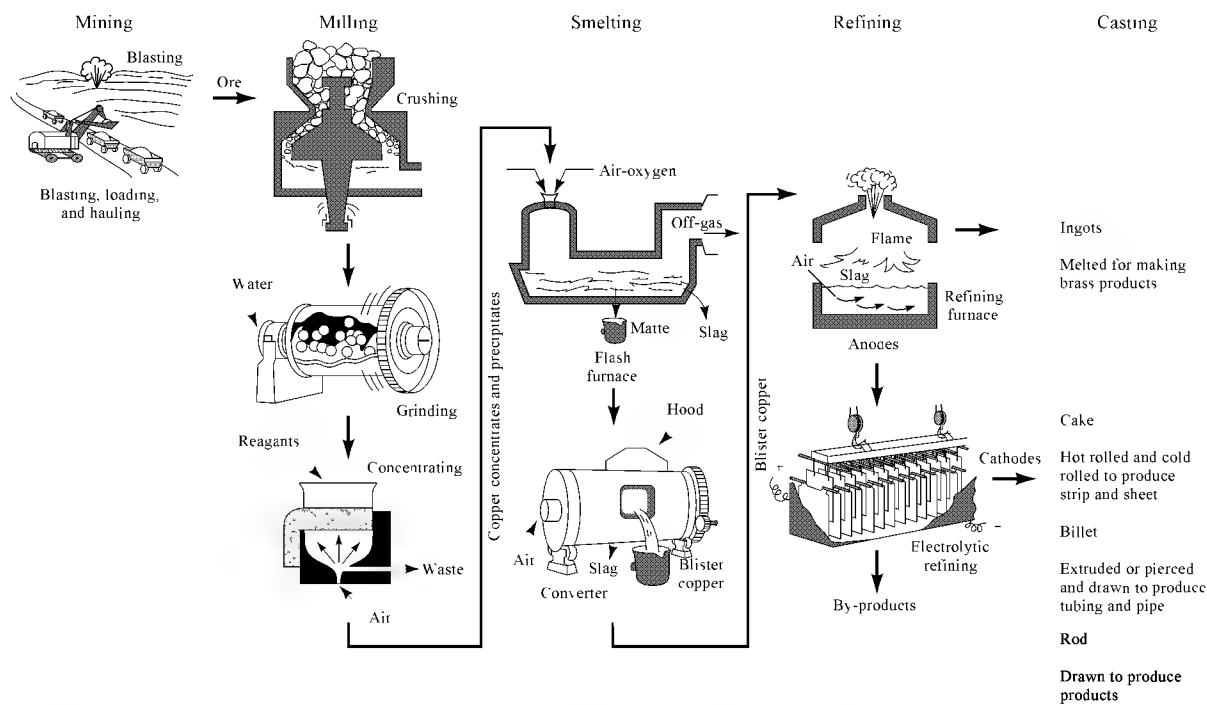


Fig. 1. Recovery of copper from sulfide ore. The residue from electrolytic refining is processed to recover gold, silver, and selenium. Courtesy of Kennecott Corp.

Hydrometallurgical processing of copper concentrates has long been suggested as an alternative approach to avoid the environmental problems intrinsic to smelting. Few hydrometallurgical plants for concentrates have been built because of high costs and problems associated with waste disposal. Hydrometallurgy is, however, firmly established in conjunction with acid leaching of ores containing oxide minerals, followed by solvent extraction and electrowinning (SX-EW) to produce pure copper for sale. In 1990, there were 14 SX-EW plants in the United States.

Mining. Copper ore may be obtained from underground or surface mines. In underground mining, the ore is removed by tunneling or shafts, whereas in surface mining the entire overburden, or waste above the ore, is removed to expose the ore. The operations in open-pit mining are drilling, blasting, loading, and hauling. Providing access for the mining and haulage equipment to remove the waste and ore leads to the characteristic inverted conical shape, with terraces or benches of open-pit mines. The trend has been toward large, open-pit mines, where the problem of declining ore grades is met by using very large capacity equipment, such as 15–25-m³ (500–900-ft³) shovels and 100–350-t haulage trucks. Most of the open-pit copper mines in the southwestern United States mine from 30,000–250,000 metric tons of ore plus waste per day.

The economics of some mines frequently depend heavily on the revenues derived from leaching the waste to recover additional copper. This waste material, which has to be removed to uncover the ore, may be hauled to specially constructed dumps, where the sulfides, the most common form of copper mineral, are oxidized and the leach solution can contact the waste material uniformly. Economics sometimes precludes optimization of dumps for leaching operations because of such factors as compaction, size distribution, and terrain (16).

Revenues from low grade copper mines are also gained from such by-products as molybdenum [7439-98-7], silver, gold, selenium [7782-49-2], and tellurium [13494-80-9].

Computer techniques are used for mine planning and design and for the optimum deployment of mining equipment. The effects of geology, haulage distances, minimum acceptable ore grade (cutoff grade), and by-product revenue are considered.

Concentration. A copper ore only rarely contains a sufficiently high percentage of copper to allow direct smelting. Ores having 1% copper or less are common. Concentrating consists of crushing and grinding to expose the copper mineralization, followed by flotation where the copper minerals are separated and recovered as a froth concentrate usually containing from 25 to 35 wt % copper. The noncopper-bearing minerals, typically silicates and carbonates, are termed gangue and are rejected as tailings, normally impounded in large ponds behind dams.

Figure 2 shows a simplified flow diagram for a typical copper concentrator. The term mill is used interchangeably with the term concentrator to describe the facility, although milling more correctly refers to just the grinding step.

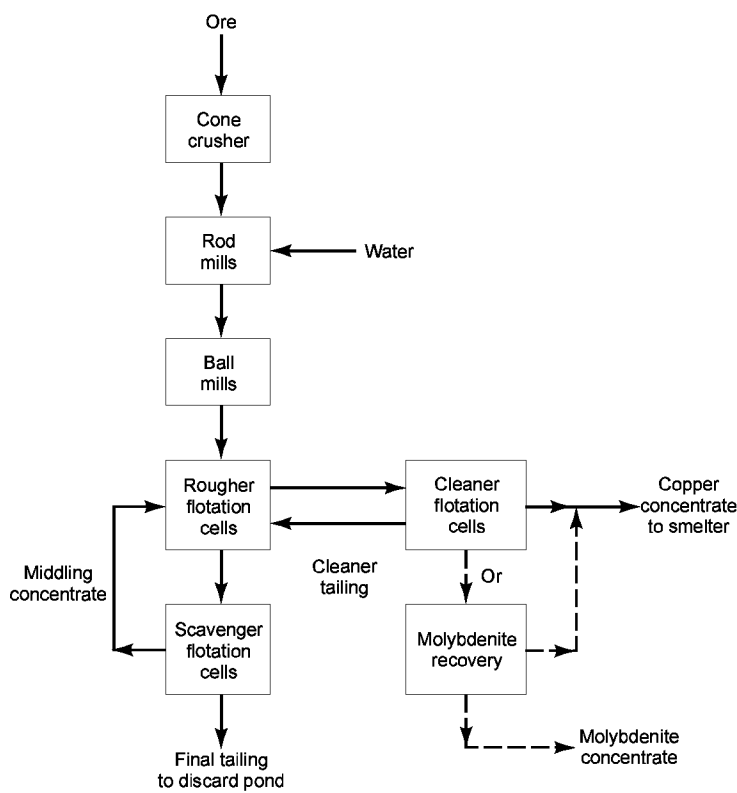


Fig. 2. Simplified flow diagram of a copper concentrator.

Crushing. Copper concentrators in the southwestern United States process ca 30,000 to 140,000 t d of low grade ore. Cone crushers are normally used for the first stages of size reduction (qv). Two or three stages of cone crushers in series are common. A cone crusher consists of an eccentrically driven grinding cone that rotates in a fixed conical bowl sloped so that there is a wedge-shaped space between the grinding cone and the bowl. The eccentrically driven grinding cone and the bowl have a vertical axis, and as ore is introduced between them it is pinched and crushed. Typical product sizes for three crushing stages in series are <20 cm, <3 cm, and <1 cm. Dry screening is used between crushing stages so that fine ore bypasses those stages in which it is finer than the crusher product.

Grinding. Further size reduction is achieved by several stages of grinding with water to produce an ore slurry. Rod mills are frequently used for the first stage, in which the ore may be reduced to <3 mm. To reduce the size of the rodmill product so that about 75% is less than 0.25 mm in size, two stages of ball mills in series may be used.

Rod and ball mills are cylindrical vessels into which the ore, water, and steel balls or steel rods, called grinding media are charged. The vessels rotate on horizontal axes to cascade the ore and grinding media and thus grind the ore. Typical ball-mill sizes range from 2.5 m in diameter and 3 m in length, driven by a 187-kW motor, up to ca 5 m in diameter and 7 m in length with a 3000-kW drive.

The product from each stage of grinding is classified by size, and the oversize material is recycled to be ground further. Wet cyclones are the most common type of classifier.

Flotation. The slurry of ground ore leaving the grinding circuit may be separated from part of the water in thickeners or may go directly to the flotation cells. The latter are rectangular tanks into which air is injected or drawn via impellers. Flotation is based on producing a water-repellent chemical film on the exposed sulfide minerals in the ground ore. The sulfide minerals collect on the surface of the air bubbles and rise to the top of the flotation cell, where they can be removed from the froth. The froth overflows the cells in collector troughs called launders.

Flotation reagents are selected to produce a stable froth and adjust the affinity of target minerals to collect in the froth. Other reagents depress the collection of minerals in the froth. The froth containing the copper minerals overflows into collection launders.

Most copper flotation plants have three banks of flotation cells, that is, rougher cells, cleaner cells, and scavenger cells (see Fig. 2). The ore slurry enters the rougher cells, where most of the copper mineralization is floated. The froth goes to the cleaner cells, where some of the gangue is carried up with the sulfides and recycled for further processing while the clean copper concentrate is dewatered and goes to the smelter. The material that does not float in the rougher cells goes to scavenger cells, where additional copper is recovered. The froth from the scavenger cells may be reground to separate gangue and mineralization before it is returned to the rougher cells.

When the ore contains a large amount of clay minerals, these form difficult to separate slimes, which hinder the recovery of the minerals (see CLAYS). The tailing from the scavenger cells can be cycloned to remove the slimes before the coarse material is floated in a tailings retreatment plant. The flotation product from the rougher cells of this plant can be reground and cleaned. This additional treatment of the tailings from the main copper flotation plant may improve the recovery of metal values by 1–3%.

The pH of the pulp to the flotation cells is carefully controlled by the addition of lime, which optimizes the action of all reagents and is used to depress pyrite. A frother, such as pine oil or a long-chain alcohol, is added to produce the froth, an important part of the flotation process. The ore minerals, coated with an oily collected layer, are hydrophobic and collect on the air bubbles; the desired minerals float while the gangue sinks. Typical collectors are xanthates, dithiophosphates, or xanthate derivatives, whereas typical depressants are calcium or sodium cyanide [143-33-9], NaCN, and lime.

Molybdenite [1309-56-4], MoS₂, normally floats with the copper sulfides. Therefore, the copper concentrate from the cleaner cells frequently has to be separated from molybdenite in a separate flotation circuit before the copper concentrate goes to the smelter. Gold, silver, selenium, and tellurium are collected with the copper concentrate.

Developments. A significant development in concentrating has been the steady increase in the size of equipment as ore grades have declined and daily throughputs have increased. Grinding mills with input power of up to 12,000 kW are used, and large flotation cells are used for efficiency of scale and energy savings. These cells may have a volume of 85 m³. The largest semiautogenous (SAG) mills are driven by 12,000-kW electric motors and process 36,000 t of ore per day.

Another innovation is the use of autogenous and semiautogenous grinding equipment. All or part of the ball charge in the ball mills is replaced by ore, so that the ore grinds itself. The mills are larger in diameter than are conventional ball mills and perform part of the crushing operation in addition to grinding. SAG ball mills may use 10–20% more energy than conventional ball mills for the same service, and fully autogenous ball mills use even more energy. Both provide operating-cost savings compared with conventional grinding.

Smelting. The term copper smelting designates the operations of melting the concentrate and extracting the copper by heat, flux, and the addition of oxygen. Copper concentrate is a mixture of the sulfides of copper, copper-iron, and iron with smaller amounts of gangue minerals. It normally contains 25–35% copper.

Sulfur is removed as sulfur dioxide by reacting the concentrate at high temperature in the presence of air and oxygen. Iron combines with silica

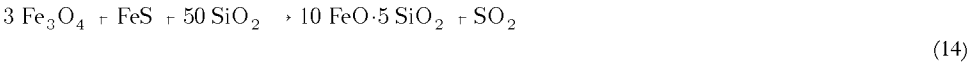
either from the gangue or the added flux and is removed together with some other impurities in the resulting slag. Part of the small amount of impurities remaining in the copper is removed by fire refining. In this step, the impurities are rejected with slag and vapor by air oxidation and slagging. Gold, silver, selenium, tellurium, and some other impurities remain with the copper to be recovered as by-products in the final purification stage, namely, electrorefining.

Reactions. When the concentrate reacts with oxygen, part of the sulfur is converted to sulfur dioxide in a concentration suitable for sulfuric acid production (see SULFURIC ACID AND SULFUR TRIOXIDE). Typical reactions occurring in the smelting vessel are



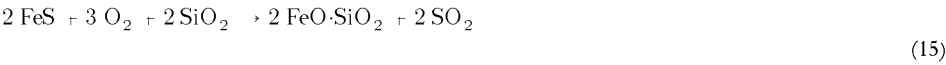
Equations 4 and 5 represent reactions occurring not only in roasting operations, that is, oxidation without melting, but also when the concentrate is not roasted before entering the smelting vessel.

When melted, copper(I) sulfide [22205-45-4], Cu₂S, and iron(II) sulfide [1317-37-9], FeS, are miscible, forming a matte typically containing 35–78% copper. Molten iron silicate and iron oxide form a separate layer, which floats on this matte and can be skimmed off in the smelting vessel to effect a partial separation of iron from copper.



The copper and copper oxide in equations 12 and 13 are from copper production via leaching or from recycled converter slag.

Because iron(II) sulfide is more readily oxidized than copper(I) sulfide, iron can be separated from copper in the matte by air oxidation. Furthermore, sulfur combines with copper rather than with iron. Hence, copper sulfide remains in the converter after the iron has been oxidized and has combined with silica to be skimmed off as a slag. Typical converting reactions of iron sulfide are equations 11, 14, and 15.



Silica is added to the converting furnace to form iron silicate slag by the reactions shown in equations 14 and 15.

Copper(I) sulfide combines with oxygen to yield copper(I) oxide [1317-39-1], Cu₂O, and sulfur dioxide. Copper(I) sulfide and copper(I) oxide react to form metallic copper and gaseous sulfur dioxide in the converters.



Fire refining, the final smelting operation, removes further impurities and adjusts the oxygen level in the copper by air oxidation followed by reduction with hydrocarbons, ammonia, or reformed gas (CO + H₂).



The thermodynamics of copper smelting are discussed in References 17 and 18. Silver and gold are quantitatively recovered with the copper throughout the smelting operations rather than being lost with the slag.

Roasting. Roasting has been largely abandoned in modern copper smelters, in which this function is combined with the smelting furnace. In older systems, the multiple-hearth roaster is a brick-lined tower having horizontal brick hearths. The concentrate is introduced at the top hearth, where rotating arms with rabble blades turn it over and move it to holes in the hearth. The concentrate is transferred successively to lower hearths and finally

leaves the bottom hearth. Air for partial oxidation of the concentrates is blown into the bottom hearth and moves countercurrently to the concentrate, leaving the top hearth with a sulfur dioxide concentration of 2–6° o.

In fluidized-bed roasters (19), the concentrate is suspended in an upward-moving air stream. The vessel is a refractory-lined steel shell with air entering through holes in a refractory-lined plate at the base. The sulfur dioxide concentration of the exit gas is 10–15° o.

The silica flux combines with iron(II) sulfide and iron(II) oxide to form slag. The fluidity of the slag, in which unwanted impurities dissolve, is controlled by the addition of limestone. Reverberatory furnaces have been largely replaced by more advanced smelting furnaces, which require lower energy input, have higher capacity, and produce higher sulfur dioxide content off-gas.

Reverberatory Smelting. Figure 3 is a simplified flow sheet for the reverberatory smelting process, showing typical material flows in metric tons per day. Concentrate, slag recycled from the converters, plus limestone flux and sometimes silica flux are charged to the reverberatory furnace. Fuel oil, natural gas, or coal burners supplement the energy obtained from the partial oxidation of the sulfides. The term calcine feed is used for concentrate that has been roasted before entering the reverberatory furnace, whereas green feed has not been roasted.

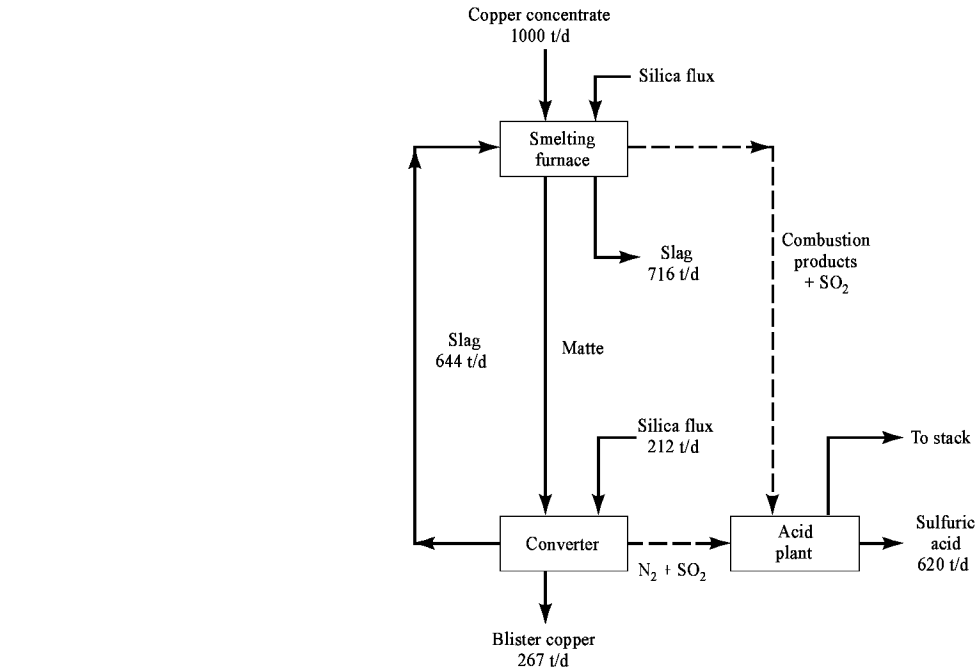


Fig. 3. Flow diagram of smelting and converting.

Converting. The most common type of copper converter is the Peirce-Smith converter, developed in the United States in 1906. The converter has changed little since that time.

Copper matte, the molten copper and iron sulfides from the smelting furnace, is charged to the copper converters using large cranes and open ladles. Silica flux is added to flux the iron in the matte. Air is blown into the charge through the tuyeres in the sides of the converters. The sulfides are oxidized to sulfur dioxide, which is collected in a hood and usually sent to sulfuric acid plants. The iron sulfide component in the matte is oxidized by blowing with air in the slag blow and combined with silica to form iron silicate slag, as shown in equation 15. In older plants, this slag may be returned to the reverberatory furnace to recover entrained and dissolved copper. Alternatively, the slag is processed by other means to recover copper, such as slow cooling and milling. After the slag has been poured off, further air in what is termed the finish blow provides oxygen to convert the copper sulfide to blister copper, containing about 99° o copper plus some sulfur, oxygen, and impurities (eq. 16–18). Figure 4 shows a typical converter. When the converter is rotated for charging and discharging, the air injection pipes, the tuyeres, are above the molten bath so that they are not blocked when the air is turned off. Partial blockages are cleared during blowing by use of mechanical tuyere punchers, steel bars that are periodically thrust through the tuyeres to dislodge any obstructions.

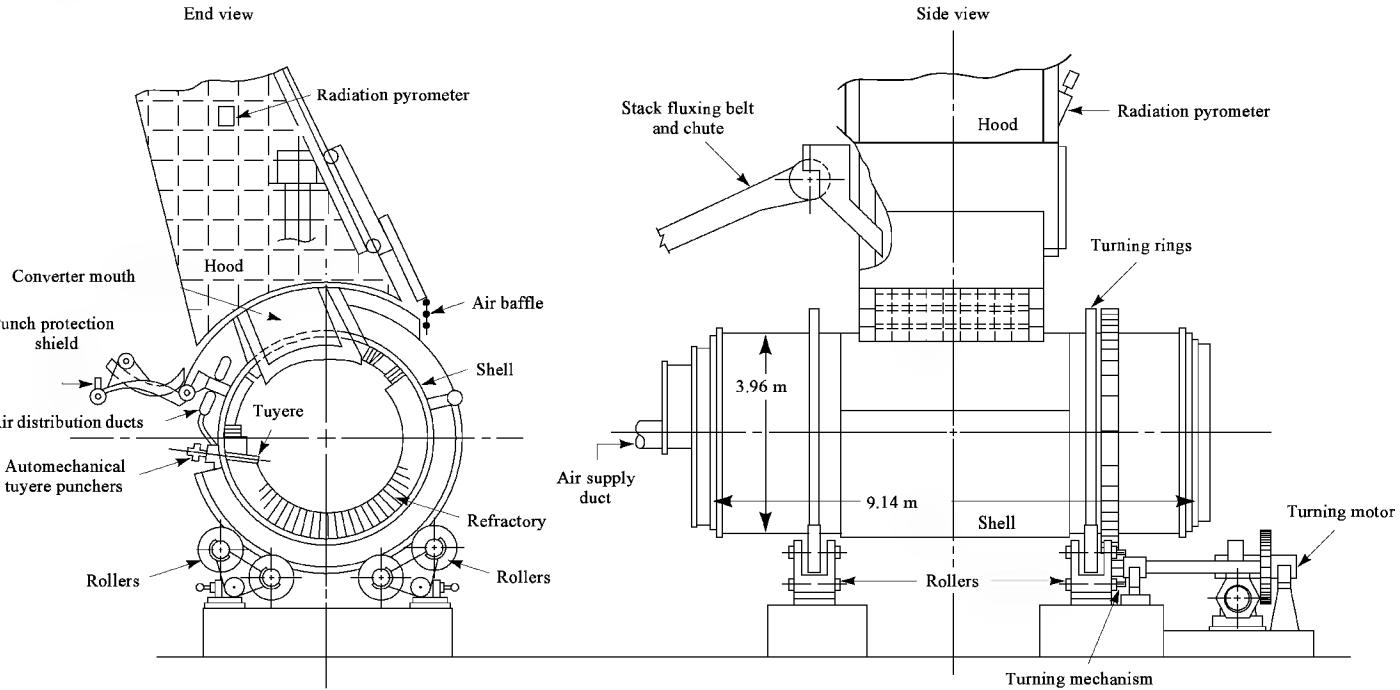


Fig. 4. The Peirce-Smith converter.

Developments. Efforts to conserve energy and comply with environmental constraints have led to more intensive smelting processes. These processes and the electric reverberatory furnace (20) are designed as direct replacements for the older reverberatory furnaces and produce matte as feed for converters. All these processes utilize oxygen-enriched air to reduce the amount of fuel required and to produce a high strength sulfur dioxide off-gas. Of these newer developments, flash smelting is the dominant process. The Outokumpu flash smelting process was developed in Finland in 1949 and is used in 38 installations around the world. The Inco flash process also involves oxygen-enriched smelting furnaces. The Mitsubishi process combines smelting and converting (21,22) and is designed for continuous production of blister copper (Fig. 5). The Noranda process produces a high grade matte, which is then treated in a converter (24,25).

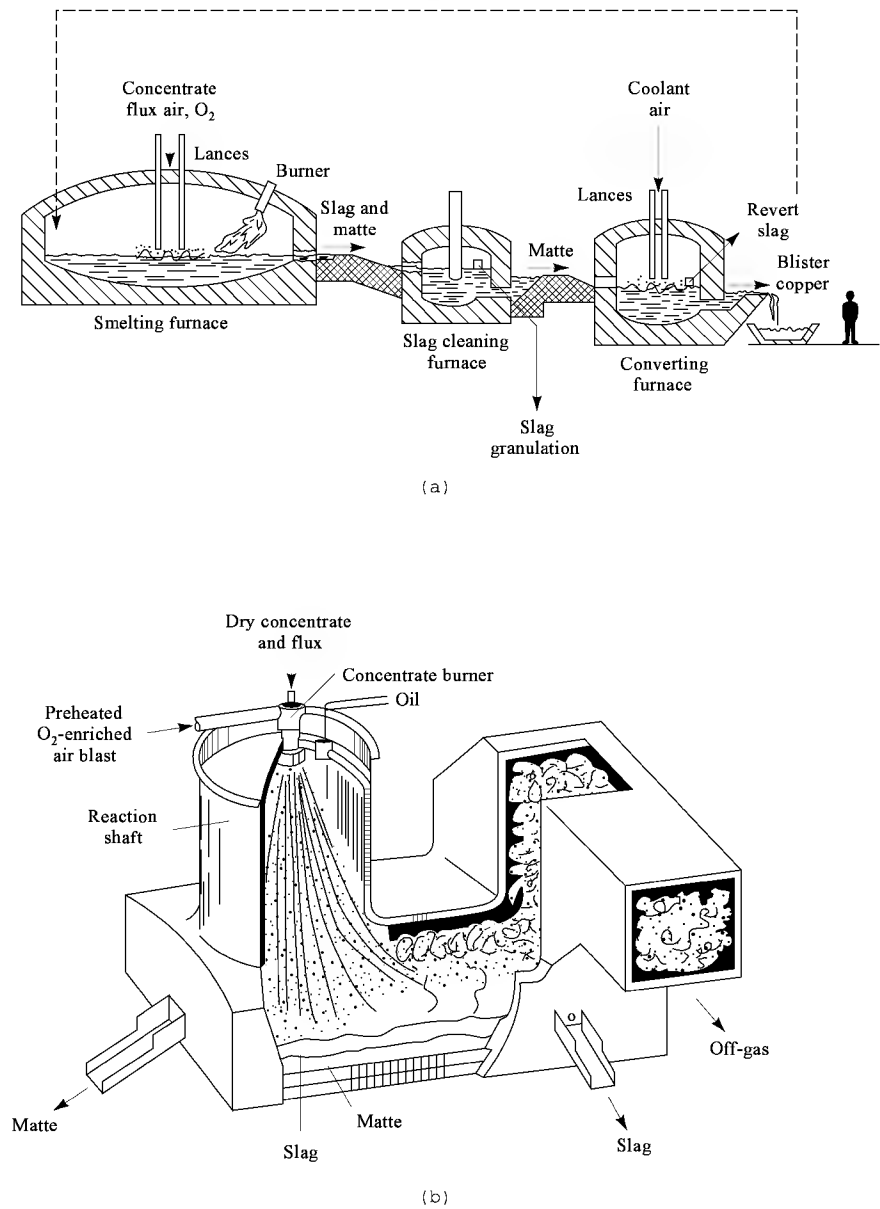


Fig. 5. Flash smelting process and equipment. (a) The Mitsubishi process. Courtesy of the Society of Mining Engineers (23). (b) Cutaway view of an Outokumpu O₂-enriched air flash furnace. Courtesy of Kennecott Corp.

In the Outokumpu, Inco, and Noranda processes, the exothermic converting reactions are carried out in part in the smelting vessel. Fuel requirements are much lower than for reverberatory smelting. The furnace slags from these processes contain appreciable concentrations of copper that must be recovered. Recovery is often accomplished in an electric slag cleaning furnace or by cooling the slag and milling it to recover the copper as a high grade copper concentrate.

A newer continuous converting technology has been developed to overcome the problems of Peirce-Smith converters. The Kennecott-Outokumpu flash converting process uses a flash furnace to treat a solidified copper matte from the smelting furnace (26). The solidified matte is produced by granulating molten matte in high pressure water sprays or by cooling the matte in molds. The matte is then ground to a suitable size and fed to a flash furnace along with oxygen-enriched air. The resulting converting reactions produce enough heat to melt the feed and produce liquid copper, very high strength sulfur dioxide gas, and a slag that is recycled to the smelting furnace. The process is continuous and avoids many of the problems associated with conventional smelting and converting. Most importantly, it eliminates the transfer of molten matte in large ladles, a practice that generates sulfur and fume emissions. This process also replaces the batch Peirce-Smith converter, which complicates emission control. A single flash converting furnace can replace four to six conventional converters. The combined flash smelting and Kennecott-Outokumpu flash converting process is shown in Figure 6.

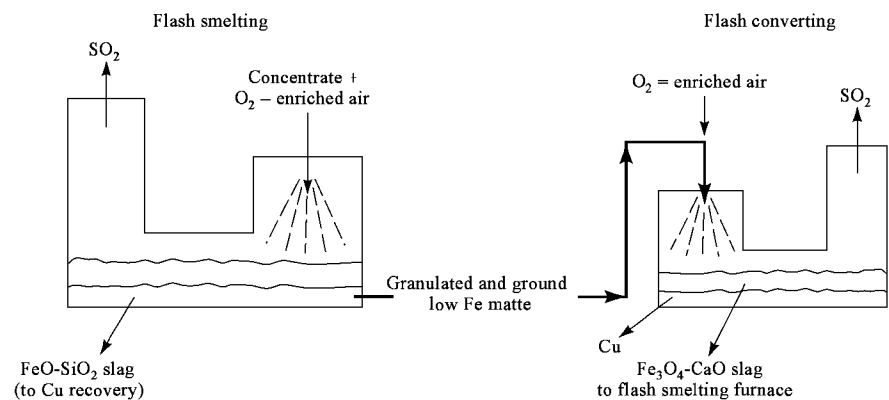


Fig. 6. Flash smelting-flash converting process for continuous production of metallic copper. Courtesy of Kennecott Corp.

Sulfur Recovery from Smelter Gases. Sulfuric acid [7664-93-9] plants are the primary means for removing sulfur dioxide from smelter off-gases. The gases from converters and roasters vary in sulfur dioxide concentration, but usually contain over 5% dioxide, which is a suitable strength for producing sulfuric acid (see SULFURIC ACID AND SULFUR TRIOXIDE). Reverberatory furnace gases from conventional equipment usually contain 0.5–2% sulfur dioxide, which is too dilute for sulfuric acid production; however, oxygen fuel burners mounted in the roof of a reverberatory furnace in the Chilean Calientes smelter (27) produce flue gases having a sulfur dioxide strength suitable for sulfuric acid production. The gases from modern smelting processes, such as flash smelting, can contain more than 25% sulfur dioxide.

The Inco flash smelting process produces a very high strength sulfur dioxide gas by using pure oxygen for smelting. Liquid sulfur dioxide is obtained upon compression. Scrubbers for removing sulfur dioxide from smelter off-gases have been under development for many years. They are widely used in Japan. The calcium sulfate (gypsum) obtained from this process is suitable feed for wallboard production (see CALCIUM COMPOUND, CALCIUM SULFATE; SULFUR REMOVAL AND RECOVERY).

Fire Refining. The impurities in blister copper obtained from converters must be reduced before the blister can be fabricated or cast into anodes to be electrolytically refined. High sulfur and oxygen levels result in excessive gas evolution during casting and uneven anode surfaces. Such anodes result in low current efficiencies and uneven cathode deposits with excessive impurities. Fire refining is essential whether the copper is to be marketed directly or electrorefined.

Fire refining adjusts the sulfur and oxygen levels in the blister copper and removes impurities as slag or volatile products. The fire-refined copper is sold for fabrication into end products, provided that the chemistry permits product specifications to be met. Some impurities, such as selenium and nickel, are not sufficiently removed by fire refining. If these impurities are detrimental to fabrication or end use, the copper must be electrorefined. Other impurities, such as gold, silver, selenium, and tellurium, are only recovered via electrorefining. Virtually all copper is electrorefined.

Blister copper is fire-refined in reverberatory or rotary furnaces similar to converters. Both types have capacities of 100–600 t. Most plants fire-refine and cast the anodes or fire-refined ingots within the smelter building so that the fire-refining furnace is supplied with molten blister copper. Solidified blister must be melted first. Next, air is blown into the copper through iron pipes or tuyeres to oxidize some impurities and remove volatile impurities. Sodium carbonate flux may be added to remove arsenic and antimony, and finally the copper is reduced by poling with green wood poles or by feeding a reducing gas such as ammonia, reformed gas, or natural gas. The degrees of oxidation and reduction are determined by removing and casting small samples of copper. An experienced operator knows whether the copper has the proper oxygen and sulfur content by the appearance of the samples.

The end product is either cast as anodes for electrolytic refining or, rarely, as ingots for sale as fire-refined copper. A horizontal casting wheel with 12–32 horizontal molds is normally used for anode casing. Using continuous casting machines, the copper is cast as a continuous strip to be cut to the required anode shape (28).

Chemistry. Fire refining of blister copper is achieved by oxidation, fluxing, and reduction. The copper is partially oxidized by blowing air into the fire-refining furnace, where the copper oxide and the impurities that oxidize preferentially react.

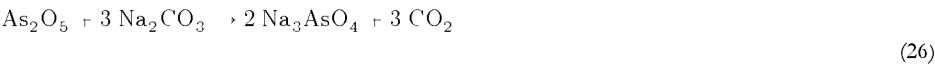
Impurities are eliminated in fire refining in the following sequence: slag, that is, oxides of iron, magnesium, aluminum, and silicon; fluxing, that is, arsenic and antimony; and vapors, that is, sulfur (as SO₂), cadmium, and zinc.

Blowing the blister copper raises the copper oxide concentration to 0.6–1.0%, which is too high for casting. The traditional reduction process is poling. In this operation, the end of a green tree trunk is inserted into the bath of molten metal in the fire-refining furnace. The resulting vigorous agitation of the bath combines with the reducing conditions from burning wood to reduce the oxygen content in the metal. The log-poling process has largely been abandoned in favor of reducing gases, however, because of cost and safety considerations.

If the fire-refined copper is to be cast into anodes for electrorefining, the oxygen content of the copper is lowered to 0.05–0.2%. If the copper is to be sold directly for fabrication, the oxygen level is adjusted to 0.03–0.05%, which is the range for tough-pitch copper. The principal reactions of fire refining are



for fluxing the equations are



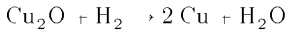
and for poling or deoxidation:



(27)



(28)



(29)

Electrorefining. Fire-refined copper is adequate for noncritical applications such as water tubing, bar stock, or ingots for alloying. Copper intended for electrical uses, however, is produced by electrorefining or sometimes electrowinning techniques, because copper having the lowest impurities is required to provide the highest electrical conductivity. Of all the industrial metals, copper is the most commonly used in its unalloyed, high purity form. Total metal impurities in the highest grade copper must be less than a few parts per million.

In the electrorefinery, anodes produced from fire-refined copper as described above are dissolved electrolytically in acidic copper sulfate and the copper is deposited on starting sheets to produce cathodes. These operations are carried out in tanks, and refineries are often known as tank houses. The cathodes are sold directly or melted and cast into a number of forms.

Anode impurities either dissolve in the electrolyte or fall to the bottom of the electrolytic cell as anode slime. These slimes contain silver, gold, selenium, and tellurium and represent a very significant value. Thus, the recovery of by-products from the anode slime is an important operation.

The basic process for electrorefining copper was developed around 1900, and all refineries use the same fundamental process. Figure 7 is a simplified flow diagram of a modern electrorefinery, consisting of four steps: (1) the anode copper is dissolved in the tank house, and pure copper is deposited electrolytically on thin starting sheets in electrolytic cells to produce cathodes; (2) the electrolyte is treated to control the concentrations of copper and impurities, for example, stripping copper from solution in electrowinning cells; (3) the anode slime, that is, the residual material collected at the bottom of the electrolytic cells during refining, is processed for recovery of precious metals, selenium, and tellurium; and (4) the refined copper cathodes are melted down and cast into commercial shapes; in addition, anodes may be cast at the refinery.

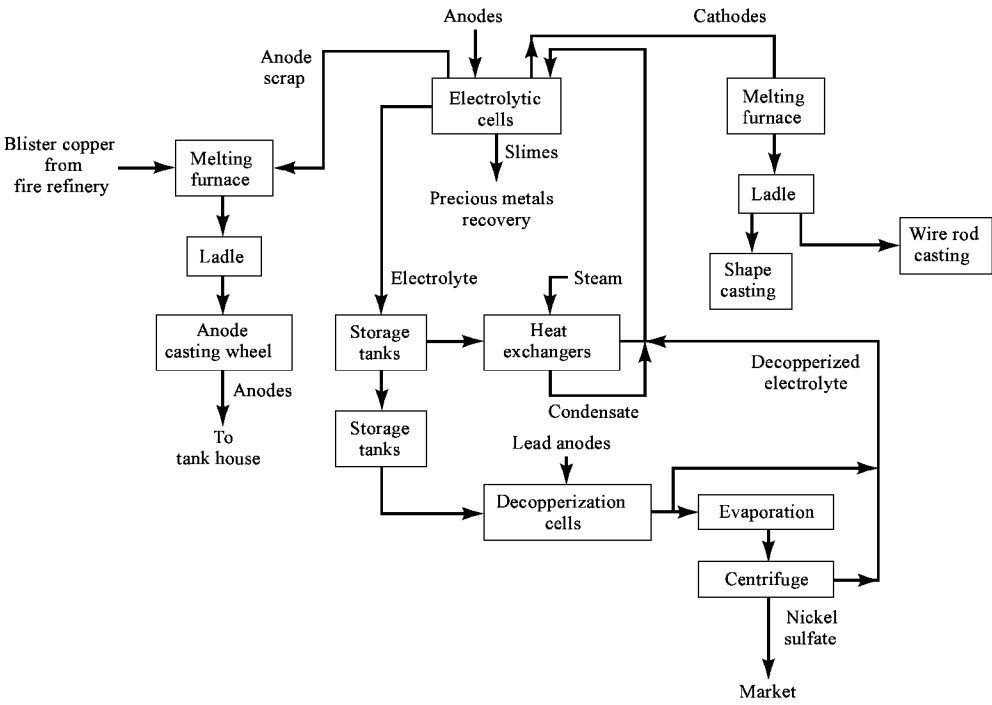


Fig. 7. Flow diagram of electrorefinery.

Electrodeposition. Electrodeposition, the most important of the unit processes in electrorefining, is performed in lead- or plastic-lined concrete cells or, more recently, in polymer-concrete electrolytic cells. A refinery having an annual production of 175,000 t might have as many as 1250 cells in the tank house. The cells are multiply connected such that anodes and cathodes are placed alternately and connected in parallel. Each cell is a separate unit and electrically connected to adjacent cells by a bus bar.

The tank house is divided into commercial and stripper sections. In the latter, one-day deposits are prepared by electrorefining anode copper onto oiled copper, stainless steel, or titanium blanks. These copper sheets are stripped from the blanks and fabricated into starter sheets for the commercial sections as starting cathodes. After 9–15 days, depending on the tank house, full-term cathodes are pulled and washed and either sent to the casting department or sold directly.

The series system was used in the past as an alternative to the multiple system. In this system only anodes were charged and a potential was maintained between the ends of each cell so that copper dissolved from one anode was plated on the adjacent anode. After a sufficient period of time, all the original anode copper was replaced by a cathodic deposit and the impurities were either in the form of anode slime or in solution. The series system demanded highly uniform anodes, a requirement that was difficult to meet with horizontal equipment.

Composition, temperature, and flow rate of the electrolyte are of great importance to the quality of the cathode deposit, and changes in any one of these parameters can have a serious effect. Storage and circulation of the electrolyte are also significant. The total volume of electrolyte in a modern tank house is typically 6000 m³ for a copper production level of ca 500 t d.

The copper contained in the electrolyte, anodes and cathodes, and the normal circulating scrap load is equal to about 10⁰ % of the annual copper production of a typical electrorefinery. Some refineries (27) have changed from the traditional 25–30 day anode cycle to a 9–14 day cycle using smaller anodes to reduce the copper inventory in spite of the higher resulting scrap load.

Current densities and cell voltages have been increased in some refineries, but the majority of refineries are still operating at 175–230 A m², copper concentrations of 30–50 g L, electrolyte temperatures of 55–65°C, and circulation rates of 10–20 L min to obtain good-quality cathodes.

In electrorefining, copper is transferred according to



The dendritic nature of the cathode deposits results in the occlusion of small but significant quantities of electrolyte and insoluble particulate matter; therefore, the dissolved impurities and suspended matter in the electrolyte are of great importance in determining the final impurity content of copper cast from cathodes. The quality of the cathode deposit is highly influenced by the use of organic additives, which modify the crystalline deposits by influencing nucleation through surface absorption or complexation. The purpose of the additive is to obtain a cathode deposit that (1) is free from deep striations or fissures that would entrap electrolyte impurities, (2) does not develop nodular growths that might cause short circuits as well as catch and entrap slimes, and (3) is not so hard as to preclude straightening bent sheets. Animal glue is most frequently used and can be modified or extended by lignin (qv) derivatives, for example, Orzan or Goulac, sulfonated oils (Avitone), casein, and thiourea, as well as other proprietary reagents.

The total impurity content of anodes used in electrorefining is usually less than 1^o o, of which oxygen is the highest, ranging from ca 0.1 to 0.25^o o. This oxygen gives copper(I) oxide, which then reacts with the acid of the electrolyte



The precipitated copper from this reaction is an important constituent of the slime that collects at the bottom of the electrolytic cells. The accumulation of copper as well as of impurities such as nickel, arsenic, antimony, and bismuth is controlled by periodic bleed-off and treatment in the electrolyte purification section.

Although some changes occur in the melting furnace, cathode impurities are usually reflected directly in the final quality of electrorefined copper. It is commonly accepted that annealability of copper is unfavorably affected by tellurium, selenium, bismuth, antimony, and arsenic, in decreasing order of adverse effect. Silver in cathodes represents a nonrecoverable loss of silver to the refiner. If the copper content of electrolyte is maintained at the normal level of 40–50 g L, and the appropriate ratio of arsenic to antimony and bismuth (29) is present, these elements do not codeposit on the cathode.

Electrolyte Purification. In electrolyte purification, copper is removed by electrowinning in cells that are similar to normal electrorefining cells, but that have lead anodes in place of copper anodes. The electrolyte is cascaded through cells similar to those used for electrorefining. As the copper is depleted, the quality of the copper deposit is degraded and the impurity levels of the resulting liberator cathodes increase. The impurity level is low, provided that the copper concentration of the electrolyte is 12–18 g L or higher. Refineries often blend these cathodes with electrorefined cathodes as charges to the melting furnaces. As copper is removed from the electrolyte by deposition on the cathodes, free acid is generated at the anode:



The final cell product contains 250–300 g L H₂SO₄ in the last stages of electrolyte purification, and antimony and bismuth precipitate, resulting in heavily contaminated cathodes that are recycled through the smelter. Arsenic and hydrogen evolved at the cathodes at these later stages react to form arsine, and hoods must be provided to collect the toxic gas.

In many refineries, nickel is the principal impurity (up to 20 g L) in the electrolyte. The nickel remains in the electrolyte as the copper is stripped out in the purification section and is recovered from the resulting acid solution by precipitation as the sulfate in evaporators.

By-Product Recovery. The anode slime contains gold, silver, platinum, palladium, selenium, and tellurium. The sulfur, selenium, and tellurium in the slimes combine with copper and silver to give precipitates (30). Some arsenic, antimony, and bismuth can also enter the slime, depending on the concentrations in the electrolyte. Other elements that may precipitate in the electrolytic cells are lead and tin, which form lead sulfate and Sn(OH)₂SO₄.

The mud or slime is collected from the bottom of the electrolytic cells and pumped to the silver refinery, where it is processed for recovery of copper, precious metals, selenium, and, in many cases, tellurium. The anode slime contains 2–20^o selenium as copper and silver selenides, whereas gold exists as the metal and in combination with tellurium. A flow diagram is shown in Figure 8.

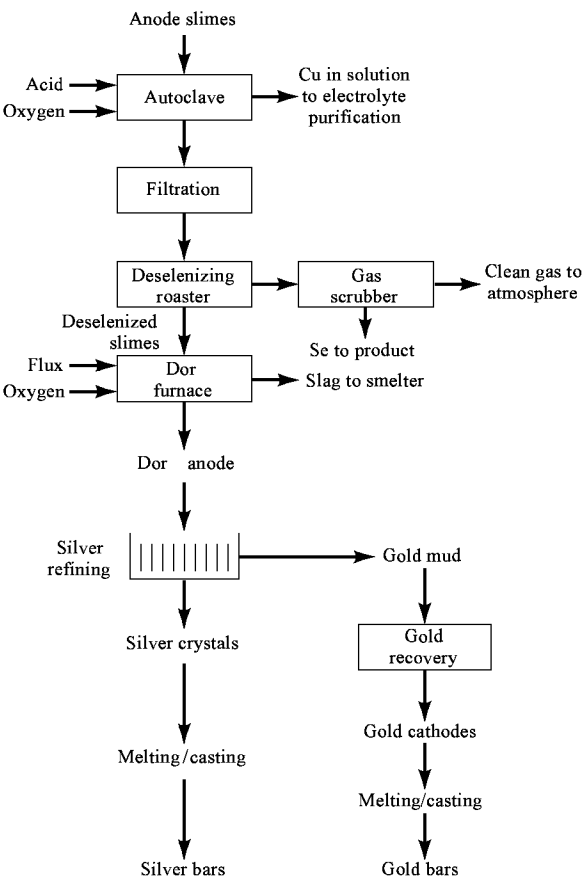
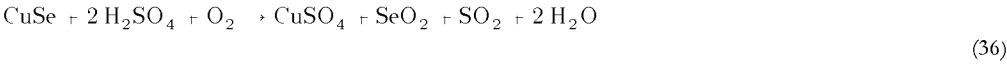
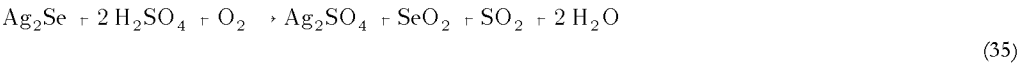


Fig. 8. Flow diagram for anode slimes treatment.

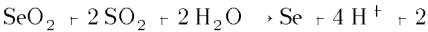
In general, the slime components are separated in silver refineries by employing combined pyrometallurgical and hydrometallurgical methods. In the first step copper is removed from the slimes in an autoclave using sulfuric acid and oxygen. The reaction is shown in equation 34.



The decopperized slimes are then roasted in the presence of sulfuric acid or sulfur dioxide and oxygen to volatilize the selenium as an oxide.



The fumes from the roaster are passed through a train of water-spray scrubbers and an electrostatic precipitator. In the scrubber, selenium dioxide [7446-08-4] reacts with sulfur dioxide (eq. 37) to produce elemental selenium, which is purified to provide a commercial product.



The roaster product is then charged to the Dorř furnace where it is melted and the resulting metal is fire-refined to eliminate the arsenates, selenates, antimonates, tellurates, and residual copper.

When the copper content in the Dorř metal has been reduced to less than 1% by fire refining, the metal is cast into anodes for electrolytic separation of silver. A typical analysis of Dorř metal is

Metal	Wt %
gold	8–9
silver	86–92
copper	0.5–1.0
palladium	0.16–0.18
platinum	0.005–0.009
lead	0.02
tellurium	0.003
selenium	0.00002

The mud or slime from the silver separation is processed to form impure gold anodes. These anodes are then electrolyzed to yield purified gold and to separate platinum and palladium for subsequent recovery (31).

Casting. Copper formerly melted in reverberatory, arc, and induction furnaces is now melted in shaft furnaces. Copper cathodes are charged into the top of a shaft furnace and natural gas burners are fired into the bottom of the shaft. The cathodes melt, and the molten copper collects in the bottom of the furnace. Shaft furnaces have the advantage of producing molten copper suitable for casting directly into rods for rolling and drawing into electrical wire. These furnaces are easy to operate and can be readily started and stopped. They use far less energy than do other melting furnaces. However, they are dependent on natural gas or propane for fuel.

In the past, wire bar, cake, billets, and a variety of semifinished products were cast in horizontal casting wheels. These have largely been replaced by continuous casting (Fig. 9).

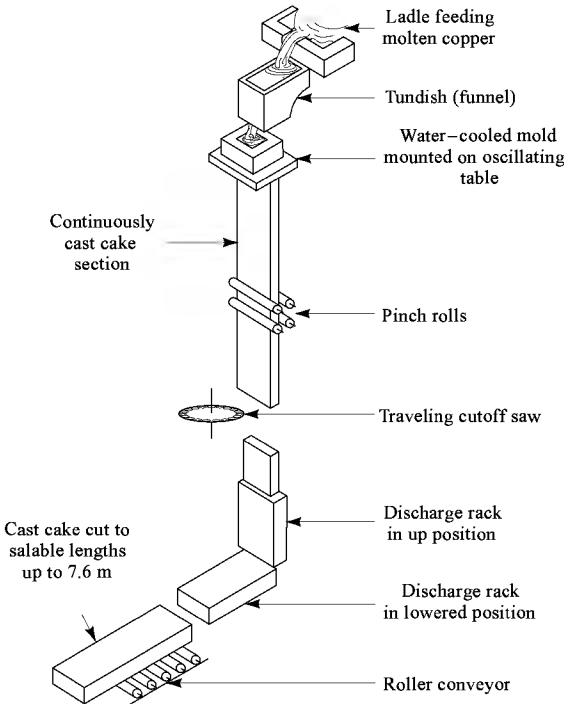


Fig. 9. Continuous casting of copper cakes.

Copper rod can be produced by four different processes without going through the intermediate wire-bar stage. In the dip-forming process developed at General Electric (32) copper is solidified onto a copper rod passed through a molten bath of copper, the Outokumpu up-cast technique. The belt-and-wheel process, based on a continuous casting concept, was developed jointly by Western Electric Co. and Southwire Co. (33). The Contirod system uses a narrow Hazeltett continuous casting machine having double belts with side dams, rather than a belt and wheel, to cast a continuous bar, which is then rolled into rod.

The rod resulting from the dip-forming process is substantially oxygen-free because the oxygen content in the molten bath must be controlled at less than 20 ppm.

Developments. Electrolytic refining requires a large capital investment, and labor costs per kilogram of copper produced are high. Most refineries have traditionally operated at current densities of about 240 A m². Thus, a tank house area of approximately 40 m² is required per ton of copper produced daily. The use of higher current densities reduces capital requirements but may impair deposition efficiency and product quality.

Current density can be increased without impairing the quality of the copper by polishing the cathode surface by brief periodic current reversals (PCR). Reversed current electrolysis, first developed for electroplating, was tested in 1952 for copper refining. Although good results were obtained, no suitable electrical equipment for current reversal was available. The thyristor-controlled silicon rectifier, introduced in the 1960s, provided a means for

industrial application of periodic current reversal. Its first large-scale development was at the Bulgarian copper refinery in Peridot. Many copper refineries have adopted reverse-current processes.

Current densities of 300–350 A m² can be used for copper electrorefining by the PCR technique without decreasing the purity of the cathodes but at the expense of a noticeable, though usually acceptable, surface roughness. The technique efficiently expands the capacity of existing plants and provides flexibility of operation for new installations. However, doubling the current density essentially doubles the power cost and increases costs for power rectifiers and electrical distribution systems. Reversing the current also increases energy consumption. This increased energy cost must be balanced against savings in capital investment and possibly labor costs.

Other approaches to increase current density without impairing cathode quality include air sparging, high velocity forced circulation of electrolyte, and the use of an abrasive belt or abrasive slurry for scrubbing the surface of the cathode.

Modernization at the Onahama Smelting and Refining Co., Ltd., utilizing continuous cast anodes and a high degree of automation, realized substantial cost savings (28). Improvements included a reduction in labor requirements and increased production per unit of tank house area. Complete mechanization led to a labor reduction of from 0.91 to 0.17 h t cathode. Inspection work was eliminated. In addition, by going to a 9-day anode-cathode cycle using continuous cast anodes, the locked-up copper inventory was decreased by 60% (27,28).

Hydrometallurgical Processes. In hydrometallurgical processes, metal values and by-products are recovered from aqueous solution by chemical or electrolytic processes. Values are solubilized by treating waste, ore, or concentrates. Leaching of copper ores in place by rain or natural streams and the subsequent recovery of copper from runoff mine water as impure cement copper have been practiced since Roman times. Most hydrometallurgical treatments have been applied to ores or overburden in which the copper was present as oxide, mixed oxide–sulfide, or native copper. Pyrometallurgical and hydrometallurgical processes are compared in Reference 34.

Hydrometallurgical processes for copper can be categorized as (1) acid extraction of copper from oxide ore; (2) oxidation and solution of sulfides in waste rock from mining, concentrator tailings, or *in situ* ore bodies; (3) dissolution of copper in concentrates to avoid conventional smelting; and (4) extraction of copper from deep-sea manganese nodules.

For operations producing 30,000 tons or less of copper annually, hydrometallurgy offers an alternative to smelting that avoids problems associated with sulfur dioxide recovery and environmental controls. Techniques include the Anaconda oxygen–ammonia leaching process, the Lake Shore roast-leach-electrowin process, and ferric chloride leaching processes for the treatment of copper sulfides. All the facilities that use these techniques encountered serious technical problems and were shut down within a few years of start-up.

Since 1965 the increased application of leaching technology has had an important impact on copper production. Table 5 shows the relative amounts of copper produced in the United States by various methods.

Table 5. U.S. Production of Copper-Bearing Ores and Recoverable Copper Content of Ores, t^a

Source and treatment process	1985		1986		1987		1988		1989	
	Gross weight	Recoverable copper	Gross weight	Recoverable copper	Gross weight	Recoverable copper	Gross weight	Recoverable copper	Gross weight	Recoverable copper
mined copper ore										
concentrated	164,029,000	905,537	170,020,000	906,072	201,434,000	991,857	222,268,000	1,113,287	230,526,000	1,126,382
leached	1,161,000	98,453	2,456,000	135,448	1,198,000	162,324	1,308,000	227,992	6,736,000	311,885 ^b
Total	165,190,000	1,003,990	172,476,000	1,041,520	202,632,000	1,154,181	223,567,000	1,341,279	237,262,000	1,438,267
copper precipitates shipped; leached from tailings, dump, and in-place material										
miscellaneous silver ore	1,004,000	3,745	552,000	2,599	275,000	1,194	464,000	2,098	538,000	2,355
lead ore	6,433,000	13,410	3,336,000	7,405		4,463	5,357,000	8,176		
other ores ^c	4,867,000	729	2,513,000	13,659	5,766,000	13,622	4,864,000	16,077	14,653,000	22,352
									0	
Grand total ^d		1,104,823		1,144,213		1,243,596		1,416,928		1,497,458

^a Ref. 7.

^b Includes electrowon copper from concentrates roast-leached.

^c Includes copper-lead ore, gold ore, gold–silver ore, lead–zinc ore, molybdenum ore, tungsten ore, zinc ore, fluorspar, flux ores, cleanup, ore shipped directly to smelters, and tailings.

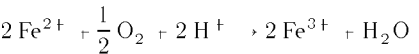
^d Data may not add to totals shown because of independent rounding.

Hydrometallurgical recovery of copper from manganese sea nodules has been studied extensively, as have combined pyrometallurgical–hydrometallurgical processes. Advancements in hydrometallurgy and leaching technology are described in References 35–37.

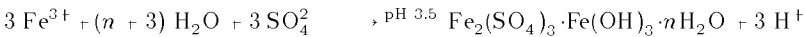
Dump Leaching. An increasing proportion of U.S. copper production is being obtained by leaching waste rock and overburden at large, open-pit copper mines. Large quantities of copper in solution have required the development of new techniques for recovering copper from solution. The nature of copper porphyry ore bodies is such that there is no distinct boundary between barren country rock and the ore body itself. Consequently, in open-pit mining of porphyry ores, considerable quantities of copper are transported to waste dumps along with the large quantities of overburden or waste rock that must be removed to reach ore-grade material suitable for treatment in flotation circuits.

Systematic attempts to maximize copper recovery from waste dumps commenced in the early 1960s. Earlier hydrometallurgical recovery of copper consisted simply of allowing normal underground water flow from mines, or water that had percolated through dumps, to pass over iron scrap in horizontal launders or troughs. Copper precipitated, and the iron passed into solution. Recirculating the solution from which the copper had been removed was not systematic. Since then it has been recognized that waste dumps represent a large potential source of copper, and a number of copper producers have increased copper recovery from waste dumps.

The following factors are important in dump leaching: (1) the role of bacteria; (2) the application of acid to prevent or delay precipitation of hydrated ferric sulfate; (3) oxidation to remove excess iron from mine water in settling pools, as shown in equations 38 and 39; (4) optimization of dump configuration for good solution distribution; and (5) availability of oxygen.



(38)



(39)

Many waste-rock or overburden disposal systems result in compacted dumps having uncontrolled distribution of fines. In such dumps, solution distribution is poor and there is little oxygen for reaction with the sulfides. Methods for managing these dumps to maximize copper recovery have been actively pursued.

Vat, Heat, or Agitated Leaching. Vat or agitated leaching usually extracts copper from oxide or mixed oxysulfide ores containing more than 0.5% acid-soluble copper. These methods are preferred over heap leaching if the ore is not porous and if crushing is necessary to permit adequate contact between the leach solution and the copper minerals. Advantages of vat over heap leaching include rapid copper recovery using a higher extraction, reduced solution losses, and higher copper content in the effluent solution. Disadvantages are higher capital and operating costs, and shorter leach-cycle times.

In-Place Leaching. Whereas *in situ* or solution mining has been defined as the in-place extraction of metals from ores located within the confines of a mine or in dumps, heaps, slag piles, or tailings ponds, in-place leaching is discussed herein as the leaching of ore without its removal from the ground. Figure 10 illustrates three solution mining techniques according to the depth of deposit.

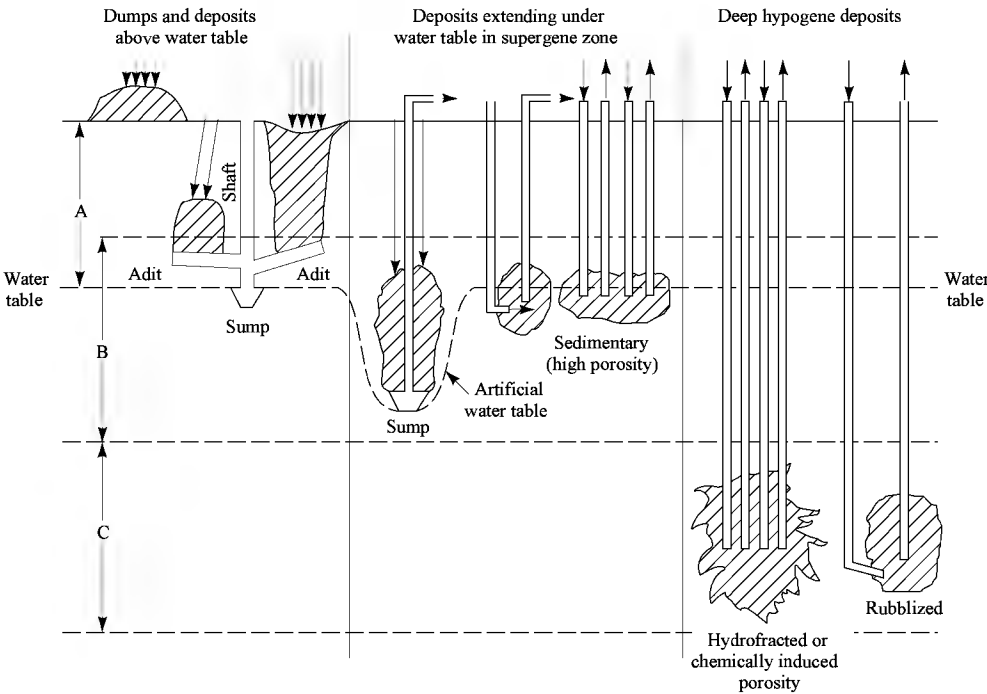


Fig. 10. Three types of solution mining: rubblized (explosives or mining), hydrofracted, or chemically induced porosity, where A represents the oxidized; B, the supergene; and C, the hypogene zones.

Courtesy of the Society on Mining Engineers (38).

Several open file reports on *in situ* copper mining have been issued, including a generic *in situ* copper mine design manual (40). Environmentally sound, low cost, *in situ* leach mining operations have been designed for small, deep, and low grade copper oxide deposits. Several procedures for *in situ* operations have been developed. One is a systematic method for assessing the commercial feasibility of the deposit (40).

Cementation. Cementation is the precipitation of copper from copper leach solutions by replacement with iron. It was formerly the most commonly used method of recovering copper from leach solutions but has been replaced by solvent extraction–electrowinning. The type of iron used in cementation is important, and the most widely used material is detinned, light-gauge, shredded scrap iron. This operation can be performed by the scrap iron cone (Kennecott Precipitation Cone) or a vibrating cementation mill that combines high copper precipitation efficiency and reduced iron consumption (41).

The fast copper precipitation rates obtainable using sponge or particulate iron promise economic and processing advantages if particulate iron becomes cost-competitive with scrap iron. Another precipitant of potential importance is shredded automobile scrap using drum-type precipitators (42).

Solvent Extraction. Solvent extraction in combination with electrowinning is the most common approach to recovering copper from acidic solutions. A variety of extractants based on aldoximes are available and show superior performance to the earlier ketoximes (43).

All commercial copper extractants selectively extract copper from weakly acidic aqueous leach solutions by the general reaction



where RH is an organic compound and RH and R₂Cu are not soluble in the aqueous solution. In the extraction reaction two protons are exchanged for the copper in the leach solution. Contacting the copper-bearing organic solution with a strongly acidic aqueous solution reverses equation 40 to allow stripping of the copper.

A flow sheet of solvent extraction–electrowinning (SX–EW) is shown in Figure 11. The organic reagent transfers the copper from dilute impure leach solutions to concentrated higher purity solutions that are appropriate feeds for electrowinning plants. The SX–EW plant has three basic functions: extracting, stripping (both in mixer settlers), and electrowinning (in the electrowinning tank house).

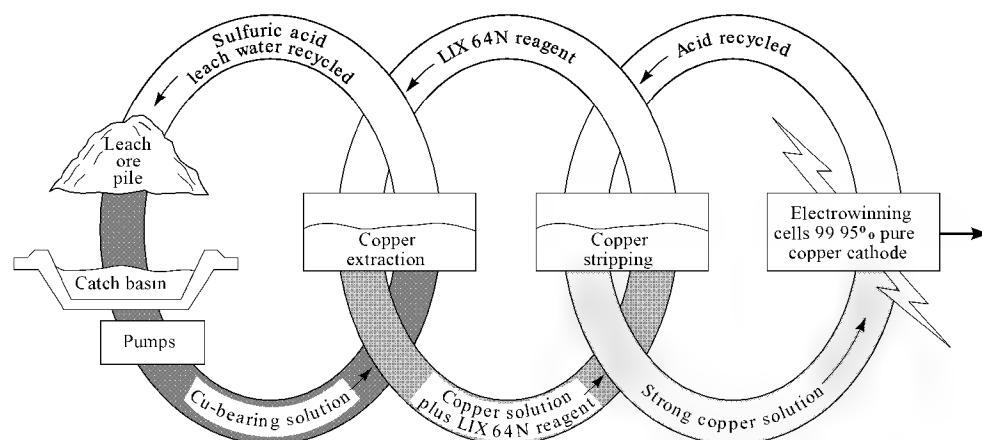


Fig. 11. Conceptual flow sheet of solvent extraction–electrowinning.

Courtesy of General Mills Chemical Incorporated.

The use of an extractant depends on loading capacity, extraction rate, pH range, and the cost of the reagent and the diluent. Loss of the extractant must be minimized because of its high cost. Organic losses to the aqueous phase are also undesirable because of the deleterious effect on cathode deposits. Advances in SX–EW processes are described in Reference 38.

An organic reagent system for solvent extraction of copper should exhibit (1) extremely low solubility in aqueous solutions; (2) high solubility in diluents such as kerosene; (3) low rates of biological or chemical degradation; (4) high copper-loading capacity; (5) rapid rate of copper extraction; (6) low cost; (7) good phase-separation characteristics; (8) ability to extract copper in the 1–7 pH range; (9) high selectivity; (10) low toxicity; and (11) low volatility at the operating temperature.

Electrowinning. Vat leaching often yields copper solutions having concentrations sufficiently high for direct electrowinning. However, high concentrations of cations other than copper and low copper concentrations make it more difficult to obtain high purity electrolytic copper by direct electrolysis of leach solutions than by electrolysis of purified solutions obtained from solvent extraction.

Increasingly more copper is being produced by electrowinning because of economics and technical advances, such as in solvent extraction processes. Certain brands obtained by SX–EW are treated as cathode quality and are used directly by wire-rod manufacturers. Whereas in 1984 100,180 t of copper was electrowon in SX–EW plants, in 1992, 439,043 t produced by SX–EQ was electrowon (7).

Copper obtained by electrowinning tends to have a high lead content, partly because of the use of lead alloy anodes. Presence of cobalt in the electrolyte or substituting calcium for antimony in the lead anodes inhibits anode corrosion. The levels of other impurities, such as arsenic, bismuth, selenium, tellurium, and silver, are usually lower in copper obtained by electrowinning from solvent-extraction solution than in copper from conventional electrorefining processes. Although the production of cathode-grade copper at the mine site is a strong incentive for using SX–EW, the fundamental benefit of SX–EW is its substantially lower operating cost.

In electrowinning, the cathode reaction is the same as for electrorefining (see eq. 31). However, because of the use of insoluble anodes, oxygen is released at the anode,



and the net reaction is



In contrast with electrorefining, there is a minimum cell voltage of ca 1.67 V, below which there is no appreciable current flow. Hence, the energy yield is only ca 0.3 kg of copper per kilowatt hour, as contrasted with about 3 kg/kWh for electrorefining.

Iron, a common contaminant of electrowinning solutions, lowers the current efficiency by dissolving copper:



The ease with which the ferrous ion can be oxidized to a ferric ion in the electrowinning cell furthers this reaction. Attack on the copper is most apparent at the solution line, where it results in corrosion of the loops supporting the cathodes, leading to dropped cathodes.

Secondary Recovery. Metal returning from the store of metal in use is referred to as old scrap, in contrast with scrap generated within the copper fabrication process, which is called new scrap (see RECYCLING). In 1990 the amount of the U.S. copper supply derived from old scrap was 24% of the total copper consumed. About 40% of old scrap is used for producing refined copper; most of the remainder is used in the production of brass and bronze ingots (see COPPER ALLOYS). About 75% of new scrap is consumed by brass mills, with most of the remainder used in the production of refined copper. Some estimates suggest that as much as 60% of the copper produced is ultimately recycled for reuse. Old scrap combined with new scrap from fabricating plants accounts for about 40% of the metallic input to domestic copper furnaces.

Effective recycling of copper scrap is an important factor in energy conservation. Processing scrap copper uses 4–40% of the energy required for primary copper production, depending on the purity of the scrap. An assessment of scrap metal recycling has been published (44). Grading of scrap follows standard definitions established by the Institute of Scrap Recycling Industries (45). The classifications are based on the original source, that is, wire, turnings, or miscellaneous material, and on the presence of contaminants.

Effluents from Production.

Gaseous. Sulfur dioxide is produced when copper and iron sulfide minerals are oxidized, an essential step in copper production. Stringent limitations have been placed on sulfur emissions by many industrialized countries. In the United States, the EPA sets rules for plant emissions, whereas OSHA is concerned with sulfur dioxide concentrations in working areas. The source standards of the EPA were established by the Clean Air Act of 1977 and amended in 1990. The U.S. copper industry has made significant investments to meet environmental regulations, resulting in a dramatic reduction in SO₂ emissions. In 1991 some 90% of these emissions were under control. U.S. copper smelters are now a minor source of SO₂ emissions and are no longer considered a significant national contributor to pollution (46).

Formation of SO₂ in the smelting operation can be controlled by replacing reverberatory furnaces with equipment that allows the formation of relatively high strength gases that can be treated to produce sulfuric acid in acid plants (23). Alternatively, the formation of SO₂ can be prevented by using hydrometallurgical techniques for copper processing.

Solid and Liquid. Copper mining operations have always been faced with a large solid waste disposal problem caused by large tonnages of waste rock or overburden with open-pit mines, tailings from the concentration step, and slag and dust piles associated with smelters. In the United States, the Solid Waste Disposal Act, amended in 1976 by the Resources Conservation and Recovery Act (RCRA), requires solid and hazardous wastes to be managed in accordance with regulations adopted by the EPA. Storage and disposal are organized to comply with government regulations.

U.S. government regulations on water discharge require the installation of water treatment plants for smelter effluents carrying significant levels of heavy metals as well as for discharges from the concentrator tailings ponds. Copper concentrators, smelters, and refineries in the United States practice maximum water recycling. Such action has required new approaches to internal water treatment and modification of flotation reagent systems to compensate for buildup in circulating loads of organic and inorganic constituents. Total recycling of water from dump-leaching operations is common practice. Actions are needed to keep leach solutions from entering groundwater systems (see GROUNDWATER MONITORING). Uncontrolled runoff from mining operations requires impoundment with appropriate treatment before release into surface systems.

Copper Sulfate. Copper sulfate [7758-98-7], CuSO₄, is produced from copper scrap, blister copper, copper precipitates, electrolytic refinery solutions, and spent electroplating solutions (see COPPER COMPOUNDS). Data from domestic producers for 1989 (7) indicated that 64% of shipments went for agricultural uses, 24% for industrial uses, including wood preservatives, and 12% for water treatment. In agriculture, copper sulfate is principally used as a fungicide for treatment of citrus and vegetable crops (see FUNGICIDES, AGRICULTURAL).

Energy. Mining uses about 20% of the total energy requirement of copper production; milling, around 40%; smelting, from 10 to 20%; and converting and refining, the remaining 20 to 30% (47). Actual requirements vary widely, depending on the mine characteristics and type of smelter. Iron (qv) mining and steel (qv) production, lead (qv) production, and zinc production (see ZINC AND ZINC ALLOYS) each uses less for a ton of product. Pollution control accounts for a large percentage of the energy demand for smelting. Energy requirements for surface and underground mining differ. The energy needed for open-pit mines is determined by the stripping ratios, mine profiles, and the distances between the mine, concentrator, and waste dumps. In many mines, about half the energy consumed is used for transportation.

Nearly all the energy used in the concentrators is for electric motors driving ball mills, crushers, pumps, and the agitators of flotation cells. About half the energy is used for grinding the ore to the proper size.

Smelting. The fuel supplied to the reverberatory furnace is in the range of 5–6 GJ t (4.7–5.7 × 10⁶ Btu t) concentrate. Steam produced in the waste heat boiler is equal to ca 60% of the energy supplied by the fuel. The additional heat recovered from the exit gases in the recuperator to preheat the combustion air is equal to ca 10% of the energy from the fuel. Hence, the heat recovered from the furnace is equal to ca 70% of the heat from the fuels.

The waste heat recovered in the boilers is in the form of steam and used either to generate electricity or to produce compressed air. The latter is used in the converter tuyeres. If electricity is generated, it is in turn used to produce compressed air for the converters. In practice, the energy made available in the waste heat boilers is almost equal to the energy required for producing the converter air.

The heat balance shows that the heat loss from the furnace walls is only ca 11% of the energy supplied by the fuel and just slightly more than the sensible heat loss with the slag. The principal heat loss is in the stack gases and is equivalent to ca 30% of the energy supplied by the fuel.

Converters. A typical heat balance for a converter operating with a reverberatory furnace shows that converting is highly exothermic. Heat losses through the converter walls are small. Heat leaving with the slag and blister copper is substantially equal to that entering with the matte. About half the heat from the converting reaction leaves with the converter off-gas, and about half is excess heat. The latter is used for melting recycled copper, such as anode scrap or frozen matte (skulls) from ladles used to transfer matte from the reverberatory furnace to the converters. When the excess heat exceeds the cooling capacity of recycled material, blister copper is cast and returned to the converters as a heat sink.

A few smelters have waste heat boilers to recover heat from converter off-gases (48). Most smelters do not have these waste heat boilers because the cyclic operation of converters results in unsteady production of steam.

Continuous Processes. Energy may be saved by partial conversion in the smelting vessel, as in continuous smelting processes such as the Noranda and Outokumpu processes. These save fuel by using part of the excess converting heat (49–51) to smelt concentrates. Furthermore, part of the energy is recovered in the off-gases passing through waste heat boilers.

Hydrometallurgical Processes. Most of the hydrometallurgical processes require slightly more total energy than is required for smelting processes. Solution pumping uses most of the required energy via dump or heap leaching. Hence, the energy consumed depends on the height of the dump or heap and the distance from the copper extraction plant.

Scrap Copper. The energy required for copper production from scrap is appreciably less than that needed production from ores. If the scrap is of low grade and has to be smelted and refined, the energy consumption is about 45 GJ t (43 × 10⁶ Btu t) of copper (52). Scrap that only requires melting and casting uses about 5 GJ t (4.7 × 10⁶ Btu t) of copper.

Outlook. The energy consumption per unit of copper production is expected to increase as copper grades decline and as facilities are added for environmental control. A significant reduction in energy consumption could be achieved by installing new facilities; however, energy cost is only a minor component of the total cost of copper production.

Economic Aspects

The United States is largely self-sufficient with respect to copper, meeting any shortfall by imports. Australia and the CIS consume most of their production on the domestic market. Japan and Western Europe import substantial quantities of copper in the form of concentrates, blister, and refined copper. World mine, smelter, and refining capacities in 1989 are given in Table 6. Copper industries in Chile, Peru, Zaire, and Zambia are nationalized.

Table 6. World Mine, Smelter, and Refinery Capacities in 1989,^a t × 10³

Continent and country	Rated capacity ^b		
	Mine	Smelter	Refinery
North America			
Canada	953	629	610
Mexico	354	285	190
United States	1,790	1,807	2,446
Total	3,097	2,721	3,246
Central and South America			
Brazil	58	160	194
Chile	1,752	1,454	1,208
Peru	443	362	272
other ^c	5	7	6
Total	2,258	1,983	1,680
Europe			
Albania	19	15	14
Austria		45	45
Belgium		172	495
Bulgaria	103	140	140
Czechoslovakia	11	25	26
Denmark		5	5
Finland	19	100	70
France		17	47

German Democratic Republic	12	45	105
Germany, Federal Republic of		389	362
Hungary		4	14
Italy		72	104
Netherlands		8	
Norway	22	35	40
Poland	465	395	432
Portugal	110	10	10
Romania	45	60	50
Spain	40	196	252
Sweden	84	110	107
Turkey	72	81	123
USSR	680	1,053	1,070
United Kingdom		75	155
Yugoslavia	142	180	165
other	3		
<i>Total</i>	<i>1,827</i>	<i>3,232</i>	<i>3,831</i>
of which			
centrally planned economies	1,335	1,737	1,837
market economy countries	492	1,495	1,994
Africa			
Botswana	26	25	
Morocco	27		
Namibia	45	60	
Zaire	779	525	250
Zambia	627	464	625
Zimbabwe	19	34	33
other ^d	3	2	3
<i>Total</i>	<i>1,747</i>	<i>1,366</i>	<i>1,074</i>
Asia			
India	75	65	47
Iran	97	70	145
Japan	22	1,212	1,196
Korea, North	15		25
Korea, Republic of	2	185	175
Malaysia	30		
Mongolia	180		
Myanmar	18		
Oman	20	22	20
People's Republic of China	400	585	475
Philippines	244	138	138
Taiwan		60	62
other	2		
<i>Total</i>	<i>1,105</i>	<i>2,337</i>	<i>2,283</i>
of which			
centrally planned economies	595	585	500
market economy countries	510	1,752	1,783
Oceania			
Australia	372	295	287
Indonesia	150		
Papua, New Guinea	344		
<i>Total</i>	<i>866</i>	<i>295</i>	<i>287</i>
<i>Total world</i>	<i>10,900</i>	<i>11,934</i>	<i>12,401</i>
of which			
centrally planned economies	1,936	2,322	2,337
market economy countries	8,964	9,612	10,064

^a Ref. 7.

^b Both primary and secondary copper.

^c Includes mine capacity of 5000 tons for Cuba, a centrally planned economy.

^d Includes mine capacity of 1000 tons for Republic of Congo (Brazzaville), a centrally planned economy.

The United States is the world's largest copper producer and in 1989 accounted for approximately 17^o % of total world production. Table 5 lists copper production in the United States for that year. Most of the copper came from 25 mines.

A detailed statistical picture of the supply and consumption of copper and copper alloys in the United States is available annually (53). The statistics trace the flow of copper from mining and scrap collection through smelting, refining, and ingot making to the wire mills, brass mills, and foundries and then on to the final markets. Table 7 summarizes U.S. uses of copper through 1990. There are strong indications that the future demand for copper will exceed production capacity.

Table 7. United States Usage Pattern, t × 10³ ^a

Usage	1985	1986	1987	1988	1989	1990
electrical	1,480	1,410	1,494	1,592	1,571	1,561
construction	343	384	395	332	327	325
machinery	150	150	132	133	131	108
transportation	86	107	88	66	65	87
ordnance	21	21	22	22	22	22
other	85	64	66	66	65	65
<i>Total</i>	<i>2,144</i>	<i>2,136</i>	<i>2,197</i>	<i>2,211</i>	<i>2,181</i>	<i>2,168</i>

^a Ref. 7.

In 1988, a comprehensive report on the technology and competitiveness of the U.S. copper industry was issued (54). This report concludes that the revitalized U.S. copper industry could compete in all but the worst foreseeable markets and that the industry's turnaround came entirely from its own efforts, with little governmental assistance. The U.S. copper industry is a world leader in smelter and refinery production, applying modern technology and measures to improve productivity.

Forecast Trends. Copper demand is forecast to grow at a 1.9^o % rate in the United States and at a 2.7^o % rate in the world between 1989 and 2000, as shown in Table 8. Whereas consumption in the United States, Europe, and Japan is expected to increase at a relatively lower rate, other countries are expected to surpass these rates as their economies develop.

Table 8. Copper Consumption^a

Year	United States	Market economy countries	World total
<i>Copper consumed, 10³ t</i>			
1950	1,337	2,502	2,774
1963	1,712	4,489	5,531
1983	2,013	7,079	8,963
1989	2,181	8,587	10,962
2000 ^b	2,700	11,100	14,680
<i>Annual growth rates, ^o %</i>			
1950–1963	1.27	3.20	3.62
1963–1989	1.00	2.50	2.67
1983–2000 ^{b, c}	1.90	2.60	2.90
1989–2000 ^b	1.95	2.43	2.70

^a Ref. 7.
^b Forecast value.
^c Forecast 1983–2000 adapted from Ref. 39

Specifications, Standards, and Quality Control

ASTM designation B224 defines the standard classifications of copper according to the method of refining and characteristics determined by the method of casting or processing (55). The accepted basic standard for electrolytic copper wire bars, cakes, slabs, billets, ingots, and ingot bars is ASTM B5. Commercial electrolytic tough-pitch (ETP) copper normally far surpasses the specifications of ASTM B5. A summary of the characteristics of a number of copper samples collected from worldwide sources is presented in Table 9. Although the copper contents are not specified, as determined by difference, the analyses would be significantly better than 99.90 as specified in ASTM B5. In addition, conductivity values are seen to be 101^o % or more in contrast with the 100^o % specified. In fact, ETP copper ranges from 99.94 to 99.96^o % copper.

Table 9. Average Impurity Levels and Physical Properties of Wire-Bar Copper Samples^{a, b}

Assay	Concentration, ppm	Standard deviation, ppm	Number of samples
<i>Impurity levels</i>			
antimony	3.41	0.55	14
arsenic	1.39	0.31	14
bismuth	0.36	0.14	15
iron	6.07	2.41	15
lead	4.08	3.06	14
nickel	3.41	1.54	14
oxygen	327.25	96.44	16
selenium	1.10	0.92	16
silver	11.19	2.93	16
sulfur	10.00	3.08	16
tellurium	1.16	1.16	16
tin	1.63	0.83	16
<i>Physical properties</i>			
density, g mL	8.48	0.15	16
conductivity, ^o % ^c	101.44	0.22	16
half-hardness temperature, °C	215.81	28.36	16

^a Ref. 7.
^b Samples from eight sources worldwide.
^c 100^o % conductivity is 0.16328 S m.

Purity. Electrolytic copper is one of the purest of the materials of commerce. The average copper content of ETP copper, for instance, is over 99.95^o %, and even the highest level of impurities other than oxygen are found only to the extent of 15–30 ppm. Up to 0.05^o % oxygen is present in the form of copper(I) oxide. Even at these low impurity levels, properties of interest to fabricators are affected in varying degree.

Electrical Conductivity. All dissolved impurities lower the conductivity of copper (56). In ETP copper the presence of 100–500 ppm oxygen decreases the conductivity slightly because of its volume effect, but more importantly, the oxygen prevents most impurities from entering precipitation into solid solution in the metal. Consequently, the deleterious effect on conductivity is reduced. In the case of antimony and cadmium, the rate of precipitation is relatively slow but some effect is observed. Impurities such as silver, arsenic, nickel, selenium, tellurium, and sulfur do not form stable oxides and affect conductivity regardless of the oxygen content. The normal tendency of selenium and tellurium to precipitate from solution is inhibited by rapid quenching, and therefore these elements have a greater effect than do the other impurities discussed above.

Figure 12 contrasts the decrease in conductivity of ETP copper with that of oxygen-free copper as impurity contents are increased. The importance of oxygen in modifying the effect of impurities on conductivity is clearly illustrated. Phosphorus, which is often used as a deoxidizer, has a pronounced effect in lowering electrical conductivity in oxygen-free copper, but little effect in the presence of excess oxygen.

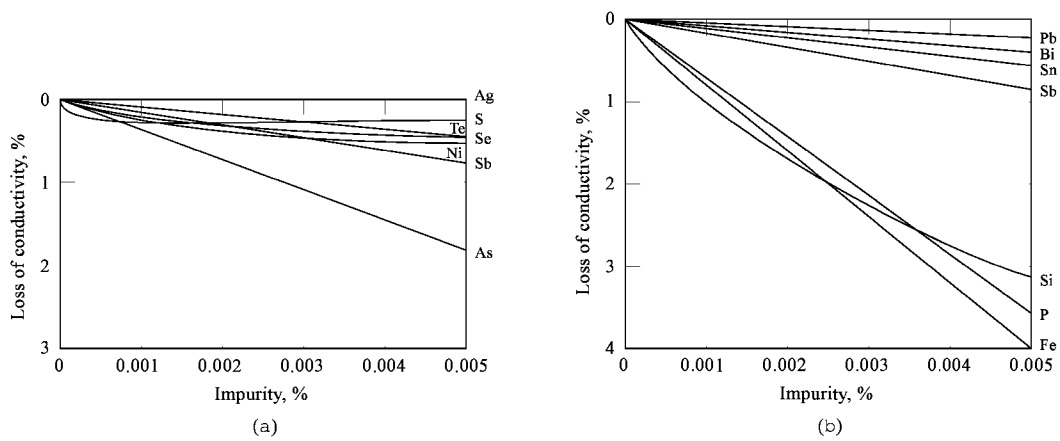


Fig. 12. Decrease in copper conductivity in the presence of the indicated impurities: (a) tough-pitch copper and (b) oxygen-free copper.

Courtesy of the American Chemical Society (56).

Fabricability. Impurities in electrolytic copper are of such low level as to have little effect on hot- or cold-working operations. The concentration of cuprous oxide affects cold-working of copper more than do minor variations of other impurities. Even so, in most applications ETP copper having 0.02–0.05% oxygen can be cold-drawn to greater than 99% reduction in area without difficulty. Oxygen-free copper and high purity copper (99.999 + %) exhibit little or no apparent limit to cold-working. Key items of importance in determining fabricability characteristics of less pure copper are the individual impurities and the corresponding solubilities and tendencies to form compounds.

Soluble impurities such as silver, gold, nickel, and arsenic do not affect hot-workability. Selenium, tellurium, sulfur, and oxygen form brittle compounds and decrease hot- and cold-workability. Sulfur contents above 0.0025% can cause problems in casting. Selenium and tellurium are sometimes added (0.5%) to improve free machining, but there is a distinct loss in fabricating, especially in cold-working properties. Bismuth and lead have limited solubility in copper and may separate as the temperature declines during hot-working of oxygen-free copper; an extremely brittle condition results. The practical limit for bismuth in oxygen-free copper is ca 0.002%. Because lead separates in globular form, ca 0.02% Pb can be tolerated in oxygen-free copper. ETP coppers containing twice these concentrations of bismuth and lead can be rolled successfully. In the production of continuous cast rod for wire production, lead cannot be tolerated at more than about 5 ppm.

Annealability. Studies of the effect of impurities on the softening of copper increased during the 1980s. Magnet wire enamels that cure at low temperatures require copper that anneals during baking at these temperatures. The enamel-coating wire must have a minimum spring-back after being wound on a form (see MAGNETIC MATERIALS).

The first studies of impurity effects were made early on in conjunction with research on the production of spectrographically pure copper (57) and involved iron and nickel in the 0.7–500 ppm range and cobalt in the 20–500 ppm range. It was concluded that in oxygen-bearing copper, that is, oxygen in the ETP range 0.02–0.05%, these impurities did not change the recrystallization temperature. In ETP copper, phosphorus had no effect on the softening temperature up to 200 ppm because of compound formation with oxygen. Arsenic (up to 5 ppm) and silver (up to 35 ppm) had no appreciable effect, and neither tin nor cadmium showed any effect because they react with oxygen. Antimony, sulfur, tellurium, and selenium were found highly effective in increasing the softening temperature of copper, subject, however, to heat treatment. The effect of antimony was studied only in the range of 18–600 ppm, an appreciably higher concentration than is normally encountered today in electrolytic copper. No combinations of impurities were investigated (57,58).

Among arsenic, bismuth, lead, antimony, and sulfur in the concentration range of 5–26 ppm, bismuth had the greatest unit effect (59). A decrease in the annealing temperature prior to cold deformation led to a decrease in the measured unit effectiveness, indicating that at low temperature bismuth is not in solid solution. Lead lowered the recrystallization temperature, provided that the samples were annealed at 700°C or lower. A precipitation reaction between lead and sulfur was proposed (60).

In a study of the springiness of copper wire, several commercial brands of copper were used (64). A strong positive relationship between bismuth concentration in the range of 2–18 ppm and springiness was found. Selenium in the range of 1–5 ppm was also very effective in increasing springiness. The effects of various impurities on springiness have been studied (62). The most deleterious impurities were found to be selenium and sulfur; silver, iron, nickel, cobalt, and oxygen had no effect. Bismuth, tellurium, antimony, and arsenic were harmful but not at the levels encountered in most electrolytic copper. The behavior of lead was erratic, sometimes appearing beneficial. The effects noted were based on the following ranges of impurities: selenium, 0.2–3 ppm; sulfur, 5–15 ppm; iron, up to 40 ppm; silver, 5–5 ppm; lead, up to 20 ppm; nickel, up to 45 ppm; and oxygen, 120–1000 ppm.

The effects of trace quantities of silver, iron, sulfur, lead, selenium, and phosphorus on the conductivity and recrystallization behavior of copper were also investigated (63). Measurements of the annealing behavior of copper rod and wire made from wire bars containing various levels and combinations of tellurium, selenium, antimony, bismuth, lead, and silver showed tellurium to have the greatest unit effect. Selenium, antimony, and bismuth also raised the softening temperature. Silver decreased the recrystallization temperature (64). The effect of lead depends on the thermal history of the copper, decreasing the softening temperature under certain annealing conditions.

Effect of Thermal History. Many of the impurities present in commercial copper are in concentrations above the solid solubility at low (eg, 300°C) temperatures. Other impurities oxidize in oxygen-bearing copper to form stable oxides at lower temperatures. Hence, because the recrystallization kinetics are influenced primarily by solute atoms in the crystal lattice, the recrystallization temperature is extremely dependent on the thermal treatment prior to cold deformation.

Heating to temperatures below 600°C has the effect of precipitating certain impurities as metallic particles and others as oxides; whereas heating to about 800°C would ensure the solution of most of the impurities in copper. Exceptions are iron, cobalt, tin, phosphorus, calcium, and perhaps zinc. Therefore, when fabricating wire it is necessary to control mill conditions for rod to achieve the lowest annealing temperature. The most important parameter is the finishing temperature, that is, the temperature prior to quenching. Allowing the hot-rolled rod to cool slowly renders many of the harmful impurities ineffectual.

Quality Control. The spectrometer is the most suitable instrument for determining most low level residual impurities. ASTM E414 is the standard method for the measurement of impurities in copper by the briquette dc-arc technique (65). In this method, the sample in the form of chips, drillings, or powder is briquetted and excited in a d-c arc opposite a high purity copper rod. Impurities in the ranges noted can be measured:

Impurity	Range, ppm	Impurity	Range, ppm
antimony	1–20	iron	5–25
arsenic	1–10	nickel	4–40
bismuth	0.1–5	tellurium	1–15
lead	3–50	tin	2–15

The specific resistance of the International Annealed Copper Standard copper is reported as $1.682 \times 10^{-6} \Omega$ at 20°C, which is defined as a

conductivity of 100^o ɔ. In performing conductivity tests, 0.48-cm (12-gauge) wire is annealed at 500°C for 20 min, given a sulfuric acid pickle, quenched, dried, and compared against a standard wire. Most commercial electrolytic copper so measured gives conductivities over 101^o ɔ.

Annealability, which is of importance in the workability of wire products, can be determined by the half-hardness temperature, residual hardness after cold-working and annealing, or spring elongation after work-hardening and annealing. The most important of these properties is the half-hardness temperature, defined as the temperature at which annealing for 1 h returns the property measured, such as hardness, tensile strength, or yield strength, to the midpoint between that for the fully worked metal and that for the fully annealed state. High purity copper, after cold-working, reaches the half-hard stage when annealed for 1 h at 140°C.

Analytical Methods Microquantities of copper are readily identified by formation of (1) the deep blue cupramine complex ion, [Cu(NH₃)₄]²⁺; (2) a red-brown precipitate in the presence of potassium ferrocyanide; or (3) a green precipitate with α-benzoin oxime. Substances containing copper, when moistened with hydrochloric acid and heated on a platinum wire, give a blue or green tinge to the flame. Numerous other reagents, chiefly organic ones, have been employed for the detection of microquantities of copper, usually by means of a spot test. Techniques such as atomic absorption analysis and x-ray fluorescence are used for control of operations and environmental measurements (see SPECTROSCOPY).

Health and Safety

Copper is required for all forms of aerobic and most forms of anaerobic life. In humans, the biological function of copper is related to the enzymatic action of specific essential copper proteins (66). Lack of these copper enzymes is considered a primary factor in cerebral degeneration, depigmentation, and arterial changes. Because of the abundance of copper in most human diets, chemically significant copper deficiency is extremely rare (67).

Accidental ingestion of large amounts of copper salts from food or beverages contaminated by copper can cause gastrointestinal disturbances, and inhalation of copper fumes can cause metal fume fever. However, no chronic copper poisoning has been reported. The human metabolic system is very efficient in promoting a discriminating copper absorption. Copper is bound to albumin in blood plasma, and large amounts can be stored in and eliminated through the liver. Therefore, although systemic effects such as hemolysis, liver damage, and renal damage have been reported after ingestion of large amounts of copper salts, recovery has usually been rapid upon treatment.

About 50^o ɔ of copper in food is absorbed, usually under equilibrium conditions, and stored in the liver and muscles. Excretion is mainly via the bile, and only a few percent of the absorbed amount is found in urine. The excretion of copper from the human body is influenced by molybdenum. A low molybdenum concentration in the diet causes a low excretion of copper, and a high intake results in a considerable increase in copper excretion (68). This copper–molybdenum relationship appears to correlate with copper deficiency symptoms in cattle. It has been suggested that, at the pH of the intestine, copper and molybdate ions react to form biologically unavailable copper molybdate (69).

Humans tolerate fairly large oral doses of copper without harmful effects, and it is used in various therapies (66). Copper sulfate is a powerful emetic and has been used clinically in the treatment of intoxications.

Copper has been employed as a bactericide, molluscicide, and fungicide for a long time and is of importance in the control of schistosomiasis (see also ANTIPARASITIC AGENTS, ANTHELMINTICS; FUNGICIDES, AGRICULTURAL). Addition of copper to lake water acts as an efficient deterrent to transmittal of the disease by eliminating snails that act as hosts for the responsible parasite. Copper is commonly utilized at ca 0.1 mg L as an algicide. In fresh water, acute toxicosis in fish is unusual if the copper concentration is below 0.025 mg L (70) (see POISONS, ECONOMIC).

The copper(II) ion appears to be toxic to marine invertebrates, and fish may be protected temporarily from copper toxicity by chelating agents (qv) such as ethylenediaminetetraacetic acid (EDTA) and nitrilotriacetic acid (NTA). Chelating agents could thus be used to detoxify copper as it passed through a critical section of a contaminated river (70). In order to eliminate the possibility of damage to ecological systems, the EPA in the United States has set limits for copper levels in discharges from nonferrous operations to lakes and streams.

In the soil, as little as 4 ppm of available copper provides the minimum requirement for most agricultural needs. Soils that are highly calcareous are likely to be deficient in copper, as are newly reclaimed peaty soils, organic chalky soils, sandy soils with a high humus content, soils that are sandy and alkaline, and highly leached soils, especially if acidic. High concentrations of nitrogen compounds can form complexes with soil copper, rendering it less available (see MINERAL NUTRIENTS). The limits of toxicological tolerance for copper in humans are so high that no problems arise in its use for parasite control on crops. Copper salts have been used for this purpose for more than 50 years (69).

Uses

The properties of wrought and cast copper and its alloys, which result in countless industrial uses, are high electrical and thermal conductivity, corrosion resistance, formability, ease of joining, machinability, low temperature strength, moderate high temperature strength, long service, recyclability, and aesthetics.

In statistical analyses of the copper, brass, and bronze industry (7), there are five primary markets: (1) building construction (41^o ɔ), including building wiring, plumbing, and heating; air conditioning (qv) and commercial refrigeration (see REFRIGERATION); builders' hardware, and architectural materials (see BUILDING MATERIALS, SURVEY); (2) electrical and electronic products (23^o ɔ), for example, power utilities, telecommunications, business electronics, and lighting and wiring devices (see also ELECTRICAL CONNECTORS; ELECTRONIC MATERIALS); (3) industrial machinery and equipment (14^o ɔ), such as in-plant equipment, industrial valves and fittings, nonelectrical instruments, off-highway vehicles, and heat exchangers (see HEAT EXCHANGE TECHNOLOGY); (4) transportation (qv) equipment (12^o ɔ), including automobiles, trucks and buses, railroads, marine vehicles, aircraft, and aerospace vehicles; and (5) consumer and general products (10^o ɔ), such as appliances, cord sets, ordnance, consumer electronics, fasteners, coinage, utensils and cutlery, and miscellaneous items.

The products for these applications are fabricated from materials from brass mills (manufacturers of copper and alloy mill products), wire mills, and foundries (manufacturers of castings). Brass mill products consist of sheet, strip, plate, rod, bar, tube, pipe, forgings, extrusions, and mechanical wire. Wire mills make electrical wire. Foundry castings consist primarily of plumbing products and industrial valves and pumps.

About 45^o ɔ of the output of U.S. brass mills is unalloyed copper and high copper alloys, chiefly in such forms as plumbing and air-conditioning tube, bus bar and other heavy gauge current-carrying flat products, and roofing sheet. Copper alloys constitute the remaining 55^o ɔ. Free-cutting brass rod, which exhibits outstanding machinability and good corrosion resistance, and brass strip, which has high strength, corrosion resistance, excellent formability, and good electrical properties, constitute more than three-fourths of the total tonnage of copper alloys produced by U.S. brass mills. Other alloy types of commercial significance include tin–bronzes (phosphor–bronzes), which are noted for excellent cold-forming behavior and strength; tin–brasses, known for outstanding corrosion resistance; copper–nickels, which are strong and particularly resistant to seawater; nickel–silvers, which combine a silvery appearance with good formability and corrosion resistance; beryllium–coppers, which provide outstanding strength when hardened; and aluminum–bronzes, which have high strength along with good resistance to oxidation, chemical attack, and mechanical abrasion.

In addition, the copper industry's market development activities have resulted in applications such as clad ship hulls, sheathing for offshore platforms, automotive electrical systems including electric vehicles, improved automobile radiators, solar energy, fire sprinkler systems, parts for fusion reactors, semiconductor lead frames, shape memory alloys, and superconducting ceramics (qv) containing copper oxides.

BIBLIOGRAPHY

"Copper" in *ECT* 1st ed., Vol. 4, pp. 391–431, by J. L. Bray, Purdue University, and H. Freiser, University of Pittsburgh; in *ECT* 2nd ed., Vol. 6, pp. 131–181, by H. Lanier, Kennecott Copper Corp.; in *ECT* 3rd ed., Vol. 6, pp. 819–869, by W. M. Tuddenham and P. A. Dougall, Kennecott Copper Corp.

1. A. J. Wilson, *Eng. Min. J.* **178**(12), 68 (Dec. 1977).
2. J. Kronsbein and R. H. Ferminhac, in A. Butts, ed., *Copper—The Science and Technology of the Metal, Its Alloys and Compounds*, ACS Monograph 122, Reinhold Publishing Corp., New York, 1954, Chapt. 21.

3. A. Sutulov, *Copper Porphyries*, University of Utah Printing Services, Salt Lake City, Utah, 1974.
4. K. Rankama and T. G. Sahama, *Geochemistry*, University of Chicago Press, Chicago, Ill., 1950.
5. E. N. Pennebaker, in Ref. 2, Chapt. 2.
6. C. Klein and C. S. Hudbut, *Manual of Mineralogy*, 20th ed., John Wiley & Sons, Inc., New York, 1985.
7. *Minerals Yearbook*, U.S. Bureau of Mines, Washington, D.C., 1989, 1992.
8. J. Jolly and D. Edelstein, *Copper, Mineral Commodity Summaries*, U.S. Bureau of Mines, Washington, D.C., 1991, p. 47.
9. U.S. Bureau of Mines, "Copper," *Mineral Facts and Problems*, Bulletin 675, 1985.
10. J. W. Laist, in M. C. Sneed, J. L. Maynard, and R. C. Brasted, eds., *Comprehensive Inorganic Chemistry*, Vol. 2, D. Van Nostrand Co. Inc., New York, 1954, p. 9.
11. R. T. Sanderson, *Chemical Periodicity*, Reinhold Publishing Corp., New York, 1960, pp. 70–99.
12. *Metals Handbook*, Vol. 2, American Society for Metals, Materials Park, Ohio, 1979.
13. Ref. 11, p. 153.
14. M. Pourbaix, *Atlas of Electrochemical Equilibria in Aqueous Solutions*, Pergamon Press, New York, 1966.
15. P. T. Gilbert, in Ref. 2, Chapt. 18.
16. S. D. Michaelson and R. E. Buckley, *Mining Engineering Handbook*, Vol. 2, Society of Mining Engineers, Salt Lake City, Utah, 1973.
17. A. Yazawa, *Can. Metall. O.* **13**(3), 1974).
18. J. M. Toguri and co-workers, *Can. Metall. O.* **3**(1), 1964).
19. MacAskill, *Eng. Min. J.* **174**, 82 (July 1973).
20. T. Nagano and T. Suzuki, in J. C. Yannopoulos and J. C. Agarwall, eds., *Extractive Metallurgy of Copper—International Symposium*, Vol. 2, AIME, New York, 1976, pp. 488–507.
21. *M.I. Process for Continuous Copper Smelting and Converting*, Mitsubishi Metal Corp., Tokyo, Japan, 1989.
22. P. Rutledge, *Eng. Min. J.* **176**(12), 88 (Dec. 1975).
23. N. J. Themelis, *Min. Eng.* **28**, 43 (Jan. 1976).
24. N. J. Themelis and co-workers, *J. Met.* **24**, 24 (Apr. 1972).
25. Ref. 19, pp. 456–487.
26. K. J. Richards, D. B. George, and L. K. Bailey, in H. Y. Sohn, D. B. George, and A. D. Zunkel, eds., *Advances in Sulfide Smelting*, Vol. 2, TMS-AIME, Warrendale, Pa., 1983, pp. 489–498.
27. J. H. Acuma, paper presented at *AIME Annual Meeting*, Atlanta, Ga., Feb. 1977.
28. H. Ikeda and Y. Matsubara, in Ref. 20, p. 588.
29. T. B. Braun, J. R. Rawling, and K. J. Richards, in Ref. 20, pp. 511–524.
30. M. Fuller-Kniepmeier and F. Pawlek, *Metall.* **1**, 36 (1963).
31. A. H. Leigh, "Precious Metal Refining Practice," in D. J. I. Evan and R. S. Shoemaker, eds., *International Symposium on Hydrometallurgy*, AIME, New York, 1973.
32. R. P. Carreker, Jr., *J. Met.* **15**, 774 (1963).
33. K. J. Kinard, *Wire* **101**, 1 (1969).
34. K. R. Suttill, *Eng. Min. J.* **191**, 30 (May 1990).
35. H. Sheffer and L. G. Evans, *Copper Leaching Practices in the Western U.S.*, Information Circular 8341, U.S. Bureau of Mines, Washington, D.C.
36. W. J. Schlitt, ed., *Leaching and Recovering Copper from As-Mined Material*, SME-AIME, New York, 1980.
37. L. A. Haas and R. Weir, eds., "Hydrometallurgy of Copper, Its Byproducts and Rarer Metals," *Proceedings of the Symposium, 1981 Annual Meeting*, SME-AIME, New York, 1983.
38. "Technology Turns Southwest Waste into Ore," *Eng. Min. J.* **191**, C41 (Jan. 1990).
39. *Mineral Facts and Problems, Bulletin 675*, U.S. Bureau of Mines, Washington, D.C., 1985.
40. D. H. Davidson, R. E. Weeks, and J. F. Edwards, *Generic In Situ Copper Mine Design Manual*, 5 vols., available from National Technical Information Service, Springfield, Va., 1988.
41. M. Esna-Ashari and co-workers, *Erzmetall.* **30**(6), 262 (1977).
42. K. C. Dean, R. D. Groves, and S. L. May, *U.S. Bur. Min. Rep. Invest.*, 7192 (1968).
43. K. R. Suttill, *Eng. Min. J.* **190**, 24 (Sept. 1989).
44. H. Spoel, *J. Met.* **42**, 38 (Apr. 1990).
45. *Recycling Copper and Brass*, Institute of Scrap Recycling Industries, Washington, D.C., 1980.
46. L. B. Rothfeld and S. W. Towle, *Eng. Min. J.* **190**, 32 (Oct. 1989); **191**, C67 (Jan. 1990).
47. C. H. Pitt and M. E. Wadsworth, *An Assessment of Energy Requirements in Proven and New Copper Processes*, report prepared for U.S. Department of Energy, contact no. EM-78-S-07-1743, Dec. 31, 1980.
48. T. Nageno and co-workers, *J. Met.* **20**, 76 (July 1968).
49. D. A. Schultz, *J. Met.* **30**, 18 (Jan. 1978).
50. H. H. Kellogg, in Ref. 28, pp. 373–415.
51. N. J. Themelis, *Min. Eng.* **27**(1), 42 (Jan. 1976).
52. H. H. Kellogg, *J. Met.* **28**, 29 (Dec. 1976).
53. *Annual Data 1991, Copper Supply & Consumption, 1970–1990*, Copper Development Association, Inc., New York, 1991.
54. Office of Technology Assessment, U.S. Congress, *Copper: Technology and Competitiveness*, Government Printing Office, Washington, D.C., stock no. 052-003-01107-7, 1988.
55. "Copper and Copper Alloys," *Annual Book of ASTM Standards*, part 2.01, American Society for Testing and Materials, Philadelphia, Pa., 1992.
56. J. S. Smart, Jr., "The Effect of Impurities in Copper" in Ref. 2, Chapt. 19.
57. J. S. Smart, Jr., and A. A. Smith, Jr., *Trans. AIME* **166**, 144 (1946).
58. V. A. Phillips and A. Phillips, *J. Inst. Met.* **81**, 185 (1952).
59. S. Lundquist and S. Carden, *Erzmetall.* **9**, 145 (1956).
60. **U.S. Pat. 2,897,107** (July 28, 1959), S. Carden and S. Lundquist (to Bolidens Gravaktiebolag).
61. S. Harper, A. R. Goreham, and J. Willmott, *J. Inst. Met.* **93**, 405 (1965).
62. K. E. Mackay and G. Armstrong-Smith, *J. Inst. Min. Metall.* **75C**, 269 (1966).
63. D. Muller, F. Pawlek, and H. Wever, *Z. Metallk.* **75C**, 269 (1966).
64. D. A. Reese and L. W. Condra, *Wire J.* **2**(7), 42 (1969).
65. "Analytical Atomic Spectroscopy; Surface Analysis," *Annual Book of ASTM Standards*, part 3.06, American Society for Testing and Materials, Philadelphia, Pa., 1992.
66. R. Osterberg, B. Sjolbert, and B. Branegard, *Final Report to INCRA*, Goteberg University, Goteberg, Sweden, Mar. 1975.
67. National Research Council Committee on Medical and Biological Effects of Environmental Pollutants, *Copper*, National Academy of Sciences, Washington, D.C., 1977.
68. *Toxicology of Metals*, Hm EOA-600 1-77-022, U.S. Environmental Protection Agency, Office of Research and Development, Health Effects Research Laboratory, Research Triangle Park, N.C., May 1977.
69. *Copper in Agriculture*, International Council for the Development of Copper, Geneva, August 1971.

70. Ref. 67, pp. 26–27.

General References

A. K. Biswas and W. G. Davenport, *Extractive Metallurgy of Copper*, Pergamon Press, New York, 1980.
Eng. Min. J., Copper USA Supplement, **191** (Jan. 1990).
I. B. Joralemon, *Copper: The Encompassing Story of Mankind's First Metal*, Howell-North Books, Berkeley, Calif., 1973.
R. F. Mikesell, *The World Copper Industry: Structure and Economic Analysis*, Johns Hopkins Press, Baltimore, Md., 1979.
R. F. Prain, *Copper: The Anatomy of an Industry*, Mining Journal Books, London, 1975.
S. D. Strauss, *Trouble in the Third Kingdom*, Mining Journal Books, London, 1986.
U.S. Bureau of Mines, annual *Minerals Yearbook*, Washington, D.C.

David B. George
Kennecott Corporation

COPPER ALLOYS

Wrought copper and wrought copper alloys,
Cast copper alloys,

WROUGHT COPPER AND WROUGHT COPPER ALLOYS

Among the metals of commercial importance, copper [7440-50-8] (qv) and its alloys are surpassed only by iron (qv) and aluminum (see ALUMINUM AND ALUMINUM ALLOYS) in worldwide consumption. Typically, copper is alloyed with other elements to provide a broad range of mechanical, physical, and chemical properties that account for widespread use. The principal characteristics of copper alloys are moderate-to-high electrical and thermal conductivities combined with good corrosion resistance, good strength, unique decorative appearance, and moderate cost (see also CORROSION AND CORROSION CONTROL). Most copper alloys are readily hot and cold formed, joined (soldered, brazed, and welded), and plated (see also SOLDERS AND BRAZING FILLER METALS; WELDING).

Chief consumers of copper and copper alloys are the building construction industry for electrical wire, tubing, builder's hardware, plumbing, and sheathing (see BUILDING MATERIALS, SURVEY); electrical and electronic products for motors, connectors, printed circuit copper foil, and leadframes (see ELECTRICAL CONNECTORS; ELECTRONIC MATERIALS; INTEGRATED CIRCUITS); and the transportation (qv) sector for radiators and wiring harnesses. Other industries include ordinance, power utilities, and coinage (see EXPLOSIVES AND PROPELLANTS; POWER GENERATION).

Alloy Designations

Copper is primarily alloyed to increase strength, however, electrical and thermal conductivities, corrosion resistance, formability, and color are also strongly affected by alloying. Elements typically added to copper are zinc, tin, nickel, iron, aluminum, silicon, chromium, and beryllium.

Copper and its alloys are classified in the United States by composition according to the United Numbering System (UNS) for metals and alloys. Wrought materials are assigned five-digit numerical designations which range from C10100 through C79999, but only the first three or sometimes four numerals are frequently used for brevity. The designations of wrought copper alloys are given in Table 1. Designations that start with numeral 8 or 9 are reserved for cast alloys (see COPPER ALLOYS, CAST COPPER ALLOYS) (1).

Table 1. UNS Designation for Copper and Copper Alloys

Alloy group	UNS designation	Principal alloy elements
coppers ^a	C10100–C15999	Ag, As, Mg, P, Zr
high coppers ^b	C16000–C19999	Cd, Be, Cr, Fe, Ni, P, Mg, Co
brasses	C20500–C28580	Zn
lead ^d ed brasses	C31200–C38590	Zn–Pb
tin brasses	C40400–C40980	Sn, Zn
phosphor bronzes	C50100–C52400	Sn–P
lead ^d ed bronzes	C53200–C54800	Sn–P–Pb
phosphorus–silver	C55180–C55284	P, Ag–P
aluminum bronze	C60600–C64400	Al, Fe, Ni, Co, Si
silicon bronze	C64700–C66100	Si, Sn
modified brass	C66400–C69950	Zn, Al, Si, Mn
cupronickels	C70100–C72950	Ni, Fe, Sn
nickel silvers	C73150–C77600	Ni–Zn
lead ^d ed nickel silvers	C78200–C79900	Ni–Zn–Pb

^a Contains a minimum of 99.3 wt % copper.

^b Contains a minimum of 96 wt % copper.

Under the UNS designation system, coppers that contain no or very small amounts of intentional alloy additives (including Ag, P, Zr, Mg, Sn), ie, 99.3 wt % minimum copper, are distinguished from the dilute copper alloys. These latter are called the high coppers and contain a minimum of 96 wt % copper. Alloys of copper are grouped according to the principal elemental addition, such as zinc, tin, nickel, aluminum, silicon, or combinations of these. Important alloys within these groupings and the associated nominal compositions are listed in Table 2.

Table 2. Nominal Compositions of Copper Alloys

Alloy group	UNS designation	Elemental composition, wt % ^a
coppers	C101 ^b	
	C110 ^c	
	C122	0.02 P
	C151	0.15 Zr
	C1572	0.4 Al ₂ O ₃
high coppers	C172	1.9 Be, 0.2 Co
	C1751	0.3 Be, 1.7 Ni
	C182	0.9 Cr
	C194	2.4 Fe, 0.03 P, 0.12 Zn
	C195	1.5 Fe, 0.8 Co, 0.6 Sn, 0.1 P
	C197	0.6 Fe, 0.2 P, 0.05 Mg
	C260	30 Zn
zinc brass	C360	35 Zn, 3 Pb
lead ^d ed brass	C425	9.5 Zn, 2.0 Sn
tin brass	C510	5.0 Sn, 0.1 P
phosphor bronze	C638	2.8 Al, 1.8 Si
aluminum bronze		

silicon bronze	C654	3.0 Si, 1.5 Sn, 0.1 Cr
	C655	3.3 Si, 0.9 Mn
	C688	22.7 Zn, 3.4 Al, 0.4 Co
	C7025	3 Ni, 0.65 Si, 0.1 Mg
	C706	10 Ni, 1.4 Fe
modified Cu–Zn cupronickel	C725	9.5 Ni, 2.3 Sn
	C729	15 Ni, 8 Sn
	C752	17 Zn, 18 Ni
nickel silver		

^a Remaining percentage is copper.

^b Contains a maximum of 0.01 wt % impurities.

^c Contains a maximum of 0.10 wt % impurities, excluding silver.

Most wrought alloys are provided in conditions that have been strengthened by various amounts of cold work or heat treatment. Cold worked tempers are the result of cold rolling or drawing by prescribed amounts of plastic deformation from the annealed condition. Alloys that respond to strengthening by heat treatment are referred to as precipitation or age hardenable. Cold worked conditions can also be thermally treated at relatively low temperatures to affect a slight decrease in strength (stress relief annealed) to benefit other properties, such as corrosion resistance and formability.

Temper. The system for designating material condition, whether the product form is strip, rod, or wire, is defined in ASTM Recommended Practice B601 (1). The ASTM system uses an alpha-numeric code for each of the standard temper designations. This system replaces the historical terminology of half hard, hard, spring, etc. Table 3 summarizes temper designations.

Table 3. Copper and Copper Alloy Temper Designations

Condition	Historical	ASTM B601
annealed	soft annealed	060
cold worked	1 4 hard	H01
	1 2 hard	H02
	3 4 hard	H03
	hard	H04
	extra hard	H06
	spring	H08
	super spring	H14
	1 2 hard RA	HR02
cold worked and relief annealed	hard RA	HR04
	spring RA	HR08
precipitation hardened alloys		
solution treated	A	TB00
solution treated + rolled	1 2 H	TD02
	H	TD04
mill hardened	AM	TM00
	1 2 HM	TM02
	HM	TM04
	XHM	TM06
	XHMS	TM08

Product Forms and Processing

The output from brass mills in the United States is split nearly equally between copper and the alloys of copper. Copper and dilute copper alloy wrought products are melted and processed from electrically refined copper so as to maintain low impurity content. Copper alloys are commonly made from either refined copper plus elemental additions or from recycled alloy scrap. Copper alloys can be readily manufactured from remelted scrap while maintaining low levels of nonalloy impurities. A greater proportion of the copper alloys used as engineering materials are recycled than are other commercial materials.

Wrought alloy product forms are varied and include plate, sheet, strip and foil, round and special cross-section bars, rod, and wire. Plate includes flat products that are above 5 mm (0.2 in.) in thickness. The generally accepted difference between rod and wire is that the former is provided in straight lengths, as intended for screw machining of parts. Wire is usually a coiled product for redrawing to a smaller diameter or for cold heading. Diameter is not a distinguishing feature between rod and wire except for C110 electrical wire that is provided as 7.9 mm (0.312 in.) diameter redraw rod.

Sheet and Strip. The manufacture of wrought copper materials starts with either semicontinuously cast slabs that are hot rolled, or cast plate that is thin enough, near 13 mm (0.5 in.), to be cold rolled directly. The surfaces of both hot rolled slabs and as-cast plate are milled to remove defects before proceeding to cold rolling and annealing operations.

Cold rolling to strengthen copper and its alloys is accomplished on a variety of mills. Four-high tandem or reversing mills, and cluster, ie, Sendzimir, mills are common. On four-high mills, and especially on cluster mills, rolling forces on the work rolls are contained to prevent uncontrolled bending and to ensure uniform thickness and flatness of the rolled product. Four-high mills are used for both in-process and finish rolling; cluster mills are most effectively used in the final rolling operation because of superior capability for maintaining tight control of thickness and shape. Cluster mills are also used for final rolling of foil having thicknesses as little as 0.013 mm (0.0005 in.).

Strengthening by cold rolling is accompanied by decreased ductility. A softening heat treatment is needed when ductility is lowered to below levels required by subsequent processing. Annealing treatments are done interchangeably as batch or continuous operations.

Batch (or bell) annealing is done within a sealed inner retort that contains a protective atmosphere. Burning natural gas is contained between the latter and an external retort. Continuous (strand or strip) annealing uses a furnace sealed at both ends to contain a protective atmosphere while permitting passage of the strip. Strand annealing is normally carried out at strip thicknesses from 0.25 mm (0.010 in.) to 3.18 mm (0.125 in.).

Neutral or reducing atmospheres are used to protect copper alloys when being annealed. These atmospheres range from nitrogen to mixtures of nitrogen and hydrogen derived from cracked ammonia (qv) to which nitrogen is often added. Products of combustion from natural gas that has been partially burned so as to contain the reducing gases, carbon monoxide and hydrogen, may be used as well (see GAS, NATURAL). Strip is usually acid cleaned following annealing.

The final processing operation for strip may be tension leveling for the purpose of improving flatness across the width and removing curvature (coil set). Strip is passed through a set of several intermeshed rolls, in serpentine fashion, to mildly deform the material in bending while under tension. The amount of bending strain is varied locally across the width of sheet to balance more highly strained wavy edges or center-of-width buckles and thereby remove these undesirable features.

Most sheet is slit to various narrow widths before shipping for use in stamping presses and other applications. Slitting is done by opposing rotary

disk knives that intersect the pass line of strip while it is moving at high line speed. The finished, slit product is shipped as flat pancake coils or is transversely wound on wide spools.

Rod and Tubular Products. Rod products having round and special cross sections are hot extruded from cast billets. Seamless tubing is either hot extruded over a mandrel held in position within the die orifice, or is formed by rotary piercing of heated billets. Tubes of both round and other cross sections can also be made at high line speed by high frequency induction welding of strip that is formed in-line before entering the welding head. Weld beads are scarfed from the internal and external surfaces, in-line with the welding operation. The tube is finally sized by cold drawing or tube reducing operations. These operations also develop a cold worked temper in the final product. Cold drawing can be done using either a fixed or floating mandrel. Alternatively, tube reducing uses semicircular grooved dies that oscillate along the tube over an internal mandrel. The latter method is capable of providing better tolerance control of diameter, wall thickness, and concentricity relative to cold drawing.

Wire. Most copper wire is cold drawn directly from continuously cast electrolytic cathode feedstock. Earlier, cast bars were hot worked to wire rods. By eliminating the hot working step, the continuous casting process offers higher quality at lower operating cost. Alloy wires, on the other hand, are hot rolled from cast bars. Semifinished copper and copper alloy wire rod is subsequently processed through drawing, annealing, and acid cleaning operations to the finished condition. A common practice, especially for very fine wire products, is die shaving of rod at an intermediate processing step to facilitate processing and provide a high quality product.

General Processing Behavior. Copper is readily strengthened by the cold working processes used to shape or form the material into the desired shape or thickness. Repeated cycles of working and annealing from the hot rolled or cast strip condition are typical for most copper alloys. Solid solution and dispersion strengthened alloys are usually annealed to their softest condition before the final working step, at which point the appropriate balance of strength and ductility is developed.

The effect of cold working by cold rolling of sheet on the yield, at 0.2% offset strain, and tensile strengths of copper sheet is shown in Figure 1. Fully annealed sheet is the starting condition, which is usually accomplished by an elevated temperature heat treatment that restores the ductility by recrystallizing the metal. As is typical for many alloys, the rolling curve exhibits an initial stage of rapid hardening followed by a lower hardening rate having a continued reduction in thickness. The initial work hardening results from the generation and entanglement of dislocations in the initial, nearly defect-free, annealed grains. Dislocation generation and motion permit the initially equiaxed grains to elongate in direction by the working operation.

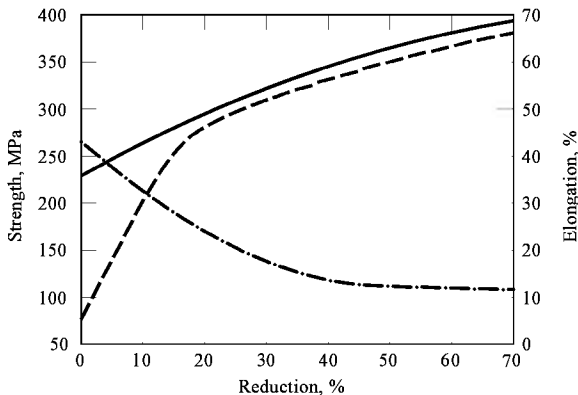


Fig. 1. The effect of cold rolling upon the tensile properties of unalloyed copper (C110): (—) represents tensile strength; (---), the 0.2% yield strength; and (-·-) the tensile elongation. To convert MPa to psi, multiply by 145.

Figure 1 also shows the decrease in tensile elongation (a common measure of ductility) that accompanies the strength increase with cold working. Whereas these particular rolling curves include up to 70% reduction in thickness, pure copper is capable of being rolled much further without fracturing.

Elements that can dissolve in copper, such as zinc, tin, and nickel for example, increase annealed strength by varying amounts depending on the element and the quantity in solution. The effect of selected solution hardening elements on tensile properties of annealed copper alloys is illustrated by the data in Table 4, where the yield strength is the stress at 0.2% offset strain in a tensile test.

Table 4. Annealed Tensile Properties of Solution Strengthened Copper Alloys

Alloy	Principal alloying element, %	0.2% Yield strength, MPa ^a	Tensile strength, MPa ^a
C110		69	255
C220	10 Zn	83	269
C260	30 Zn	145	366
C510	5 Sn	164	352
C521	8 Sn	200	414
C706	10 Ni	131	324
C752	18 Ni + 17 Zn	172	400

^a To convert MPa to psi, multiply by 145.

Solid solution hardened alloys also show different cold work hardening response relative to pure copper and each other as illustrated in Figure 2. These differences originate from the effects that the dissolved atom has on microplastic processes that occur during cold working. Finer grain sizes generally translate into stronger materials. Note that for alloy C260, finer grain size has higher strength at corresponding cold reduction. A higher quality, smooth surface that requires no buffing after forming operations is a further benefit from a finegrained microstructure.

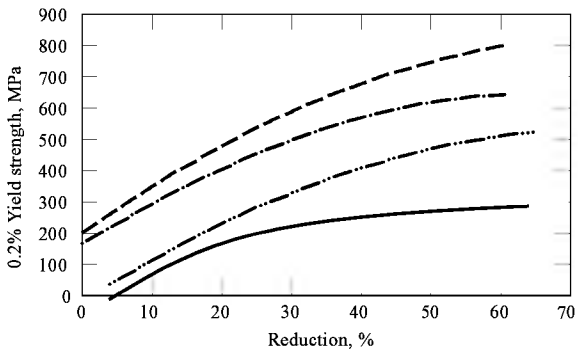


Fig. 2. Comparison of alloy cold rolling behaviors: (—) represents unalloyed copper (C110) having a grain size of 25 μm; ([|2tb2h|]) and (---), copper–30° zinc (C260) of 50 μm and 15 μm grain size, respectively; and (---), copper–8° tin (C521) having a 15 μm grain size. To convert MPa to psi, multiply by 145.

The annealing response of cold worked copper is illustrated by Figure 3 in terms of changes in tensile properties determined at room temperature after the elevated temperature anneal. Also important is the grain size obtained from each annealing temperature because this microstructural characteristic influences properties. In general, more uniform and finer grain size is promoted by increasing the amount of cold work before annealing, and by annealing at the lower temperature that becomes possible after more cold work. Many products are offered in the fully soft or annealed condition to provide maximum available formability.

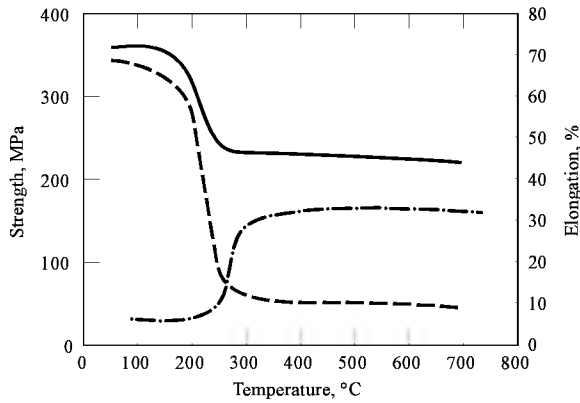


Fig. 3. Changes in tensile properties of cold rolled copper (C110), after 50° reduction in thickness, that attend annealing at each of the temperatures shown for one hour: (—) represents tensile strength; (---), the 0.2° yield strength; and (---), the tensile elongation. To convert MPa to psi, multiply by 145.

The annealing curves for several copper alloys are compared to copper in Figure 4. Solute additions affect the annealing response, as they do for cold working, by interaction with microplastic processes that account for softening as dislocations and subgrains are eliminated during the anneal.

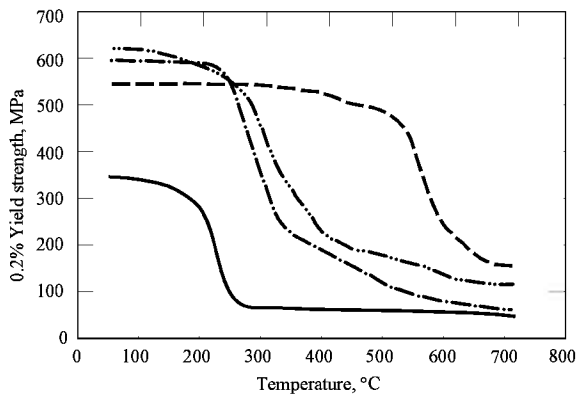


Fig. 4. Comparison of the annealing response of cold rolled (—) copper (C110) and cold rolled alloys; ([|2tb2h|]) copper–30° zinc (C260), (---) copper–5° tin (C510), and (---) copper–10° nickel (C706). The condition prior to annealing was after 50° cold reduction in thickness from the annealed-soft condition. The time of annealing at each of the temperatures is one hour. To convert MPa to psi, multiply by 145.

Alloying for Strengthening

Copper alloys can also be grouped according to how the principal elemental additions affect properties. This grouping depends primarily on whether the additions that dissolve in liquid copper can form discrete second phases during either melting casting or in-process thermal treatment. Alloy constitution that relates to limits of solid solubility and equilibrium phases that form in binary and ternary combinations with copper are found in the literature (2,3).

Solid Solution Alloys. Copper dissolves other elements to varying degrees to produce a single-phase alloy that is strengthened relative to unalloyed copper. The contribution to strengthening from an element depends on the amount in solution and by its particular physical characteristics such as atom size and valency. Tin, silicon, and aluminum show the highest strengthening efficiency of the common solute additives, whereas nickel and zinc are the least efficient (4,5). Many alloys are ternary combinations of these elements. The limiting factor in their alloy range is the extent to which these elements, either singly or in combination, remain dissolved in the copper parent during processing.

Alloys containing zinc, tin, aluminum, and nickel represent most of the commercial alloys, namely, the brasses, phosphor bronzes, cupronickels and nickel silvers (see Tables 1 and 2). The tensile properties of selected solid solution strengthened alloys, in the annealed condition, are listed in Table 4 (6). These alloys generally provide good formability, together with varying degrees of corrosion and oxidation resistance. Alloying for solution strengthening is accompanied by lowered electrical conductivity. Because of high alloy content, these solution strengthened alloys tend to have conductivities that are typically less than half that of unalloyed copper.

Dispersed Phase Alloys. The presence of finely dispersed second-phase particles in copper alloys contributes to strength, through refined grain size and increased response to hardening from cold working. These phases are developed in many commercially important copper alloys in several ways. A dispersion of fine particles can be incorporated into the alloy through thermomechanical processing where the alloy content is above the solid-state solubility limit. Precipitation and coarsening of the excess solute by an in-process anneal is used in high copper alloys, such as C194 and C195, to form iron or iron–cobalt dispersions. These dispersed phases enhance the work hardening response relative to unalloyed copper to produce high strength while maintaining reasonably good conductivity.

Solid solution alloys based on Cu–Zn and Cu–Al have also been modified by additions that react to form dispersions of intermetallic phases. Addition of aluminum and cobalt to a zinc-containing alloy (C688) and cobalt and silicon to an aluminum-containing alloy (C638) are two examples. Fine-grain sizes below ten μm and smaller amounts of cold work to develop strength combine to produce highly formable alloys that offer the highest strengths available from nonage-hardenable alloys.

Another means by which dispersed phases are produced is from those that separate from the melt during casting. These elements include iron

phosphides, as in C194 and C195, and cobalt beryllide, as in C172.

Powder metallurgy is used to incorporate second phases into copper (see METALLURGY, POWDER). An example is the manufacture of oxide dispersed alloys like C15720. Aluminum oxide powder, that otherwise does not dissolve in copper, is incorporated in C15720 by mixing powders of copper, copper oxide particles, and a dilute copper–aluminum alloy. Hot extrusion is used to consolidate the mixture. Subsequent heat treatment allows copper oxide to react with dissolved aluminum in the alloy powder to form uniformly dispersed submicrometer-sized aluminum oxide particles within the latter. It is this ultrafine oxide that is principally responsible for the alloy’s notable resistance to softening during subsequent high temperature exposure.

Precipitation (Age) Hardening Alloys. Only a few copper alloys are capable of responding to precipitation or age hardening (7). Those that do have the constitutional characteristics of being single-phase (solid solution) at elevated temperatures and are able to develop into two or more phases at lower temperatures that are capable of resisting plastic deformation. The copper alloy systems of commercial importance are based on individual additions of Be, Cr, or Ni + X where X = Al, Sn, Si, and Zr.

Processing involves heat treating the alloys at a sufficiently high temperature to dissolve solute constituents, followed by rapid cooling to near room temperature to retain this solid solution. A second lower temperature aging anneal to form the hardening phase particles is the final step. Cold work may be introduced between the two anneals to promote aging response and also add to final strength.

The preferred precipitation structure is one with uniformly dispersed particles within the grains of the alloy. Alloy composition and thermal treatments are chosen to achieve this structure through control of intermediate, metastable precipitate phases. The equilibrium, thermodynamically stable precipitate phases are generally coarse and provide little strengthening. When the latter are formed at grain boundaries, formability of the material is significantly decreased. This grain boundary precipitation is avoided for strip that is intended to be formed into parts after having been previously hardened by the mill.

Special Addition Alloys. The most notable of the special additives to copper alloys are those added to enhance machinability. Lead, tellurium, selenium, and sulfur are within this group of additives. Because of increasing concern over lead toxicity, interest has centered on use of bismuth which is nearly as effective as lead for improving machinability. The alloys that contain such additives are limited because of the difficulty they cause to hot and cold working. High zinc brasses such as C360, which undergo a change in crystal structure at elevated temperature, are favored because of the ability to accommodate lead and bismuth additions in processing. Unlike lead, tellurium and sulfur are added only to copper, as C145 and C147, respectively, and not usually to copper alloys.

Other special additions are used to deoxidize copper. Such alloys may be preferred in applications where embrittlement by hydrogen through reaction with internally dispersed copper oxide particles is a concern, such as in C110. The most common deoxidized copper is C122, in which phosphorus reacts with copper oxide to form phosphorus pentoxide that can be slagged from the copper while molten.

Many elemental additions to copper for strengthening and other properties also deoxidize the alloy. A side benefit of such additions is elimination of susceptibility to hydrogen embrittlement. Such deoxidizing additions include beryllium, aluminum, silicon, chromium, zirconium, and magnesium.

Properties

Strength. Tensile properties and electrical conductivities of selected copper alloys having commercial importance are listed in Table 5. The principal source of strengthening and the individual product forms in which each alloy is usually available are also identified.

Table 5. Properties of Copper Alloys

UNS designation	Alloy type ^a	Product forms ^b	Electrical conductivity, %IACS ^c	Modulus of elasticity, GPa ^d	Temper	Tensile strength, MPa ^e	Yield strength (0.2% offset), MPa ^e	Elongation, %
C110	pure	S, R, W, T	100	117	060	221	59	35
C151	D	S, R, W	95	117	H04	330	317	4
C1572	D	R, W	89	105	CW 97%	605	580	5
C172	P	S, R, W, T	20	128	TM04	980	860	12
C1751	P	S, R	45	130	AT	845	740	12
C182	P	S, R, W	80	130	TH04	460	405	14
C194	D	S	65	117	H08	503	482	2
C260	S	S, R, W, T	28	110	H08	662	600	2
C360	SA	S, R	26	97	H02	470	360	18
C425	S	S	28	110	H04	524	503	8
C510	S	S, R, W	15	110	H08	696	682	2
C638	SA	S	10	117	H04	834	750	5
C688	SA	S	18	117	H08	880	800	2
C655	S	S, R, W	7	105	H04	650	400	8
C706	S	S, R, T	9	124	H04	530	517	1
C7025	P	S	40	130	TM03	690	655	5
C725	S	S, R, W	11	138	H08	641	620	1
C729	P	S, W	7	127	TM04	860	790	15
C752	S	S, R, W	6	124	H08	655	641	1

^a S is solution strengthened; D, dispersed phase; P, precipitation strengthened; SA, special addition.

^b S is sheet and strip; R, rod and shapes; T, tube; and W, wire.

^c IACS – International Annealed Copper Standard. Pure copper has a value of 100.

^d To convert GPa to psi, multiply by 145,000.

^e To convert MPa to psi, multiply by 145.

Table 6 illustrates the increase in strength and the accompanying decrease in conductivity that derives from alloying of copper. This trend is clearly apparent among solid solution strengthened alloys, namely those that contain zinc, tin, nickel, and zinc plus nickel as their principal alloying constituents. Notable exceptions to the association of low conductivity with high strength are some alloys that are precipitation strengthened, namely C1751, C182, and C7025. For the latter group, both high strength and moderate conductivity are possible.

Table 6. Conductivity vs Strength Among Copper Alloys

UNS designation	Principal alloy element, %	Conductivity, annealed, %IACS	0.2% Yield strength hard temper, MPa ^a
C110		100	310
C210	5 Zn	56	365
C230	15 Zn	37	420

C260	30 Zn	28	495
C505	1.25 Sn	48	425
C510	5 Sn	15	560
C521	8 Sn	13	590
C706	10 Ni	9	520
C715	30 Ni	4	540
C1751	0.4 Be, 1.8 Ni	45	740
C182	1 Cr	80	400
C7025	3 Ni, 0.65 Si, 0.15 Mg	40	690

^a To convert MPa to psi, multiply by 145.

Alloy selection is not made from only consideration of strength and conductivity. For example, the cupronickels have about the same strength as do copper–zinc brasses, and also have much lower conductivity. However, the corrosion resistance of the cupronickels far exceeds that of brass and is worth the higher cost if needed in the application. Similar trade-offs exist between these properties and formability, softening resistance, and other properties.

Electrical–Thermal Conductivities. Electrical conductivities of alloys (Table 5) are often expressed as a percentage relative to an International Annealed Copper Standard (IACS), ie, units of % IACS, where the value of 100 % IACS is assigned to pure copper having a measured resistivity value of 0.017241 Ω mm² m. The measurement of resistivity and its conversion to % IACS is covered under ASTM B193 (8).

Copper has a high electrical conductivity that is second only to that of silver. The conductivity of silver in % IACS units is 108; gold, 73; aluminum, 64; and iron, 18. Wrought copper having a conductivity near 102% IACS is not uncommon because of improvements in refining practices since the standard was first established.

Electrical conductivity of copper is affected by temperature, alloy additions and impurities, and cold work (9–12). Relative to temperature, the electrical conductivity of annealed copper falls from 100 % IACS at room temperature to 65 % IACS at 150°C. Alloying invariably decreases conductivity. Cold work also decreases electrical conductivity as more and more dislocation and microstructural defects are incorporated into the annealed grains. These defects interfere with the passage of conduction electrons. Conductivity decreases by about 3–5% IACS for pure copper when cold worked 75% reduction in area. The conductivity of alloys is also affected to about the same degree by cold work.

Copper and its alloys also have relatively good thermal conductivity, which accounts for their application where heat removal is important, such as for heat sinks, condensers, and heat exchanger tubes (see HEAT EXCHANGE TECHNOLOGY). Thermal conductivity and electrical conductivity depend similarly on composition primarily because the conduction electrons carry some of the thermal energy.

To a good approximation, thermal conductivity at room temperature is linearly related to electrical conductivity through the Wiedemann-Franz rule. This relationship is dependent on temperature, however, because the temperature variations of the thermal and the electrical conductivities are not the same. At temperatures above room temperature, thermal conductivity of pure copper decreases more slowly than does electrical conductivity. For many copper alloys the thermal conductivity increases, whereas electrical conductivity decreases with temperature above ambient. The relationship at room temperature between thermal and electrical conductivity for moderate to high conductivity alloys is illustrated in Figure 5.

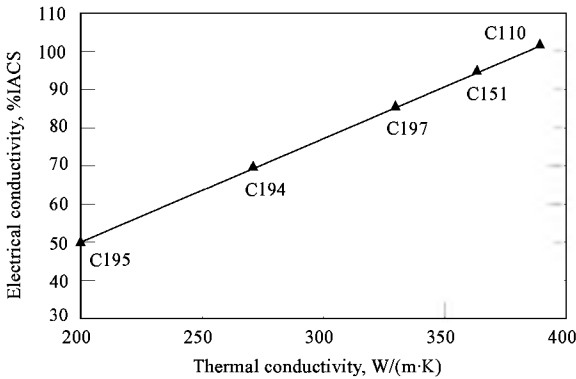


Fig. 5. The Wiedemann-Franz relationship at 20°C between electrical and thermal conductivities of copper alloys having moderate to high conductivities.

Formability.

Copper and most of the wrought alloys are readily formed by bending, drawing, upset forging, stamping, and coining (13–15). The maximum formability condition is the fully soft or annealed condition. When additional strength or hardness is desired in the final part, the forming step is done starting with a cold worked temper or the part is not annealed between forming steps. Cold forming operations always work-harden alloys and in many cases sufficient strengthening is produced.

A variety of tests have been established to indicate whether a specific forming operation can be safely accomplished without failure. Three of the most commonly used tests are bend testing around a mandrel, limiting draw ratio testing, and bulge height formability testing. Temper, strip thickness, grain size, crystallographic texture, and alloy composition are important in these tests. An alloy in the annealed temper that can be formed into a particular shape may not be able to be similarly formed starting from a work hardened temper.

Bending. The smallest radius over which strip of a particular alloy can be formed without failing is important in the selection of materials for a given application. The industry tests formability using samples cut from strip material to rank materials and thereby indicate whether a particular alloy temper is suited for an application (15). Performance of the material in actual stamping is the best final judge of suitability.

In this test, a rectangular strip sample is formed around a die with a precisely machined edge of a known radius. The sample is formed by 90° or 180° about an axis in the plane of the sheet. The outer surface of the bend is inspected for cracking or unacceptably deep surface rumpling for the particular application. The minimum bending radius (MBR) about which strip can be successfully formed depends on the strip thickness, t, and is reported with the test thickness specified or normalized with regard to thickness as MBR t. Better formability is indicated by smaller MBR t values.

Formability is not the same for all orientations of rolled temper alloy strip because of crystallographic texturing or mechanical fibering. The amount of directionality relative to the rolling direction depends on the amount of final cold reduction to temper, the intermediate processing history, and the specific alloy. Typically, bend formability is better around an axis that is perpendicular to the rolling direction relative to one that parallels the rolling direction. The industry has termed these goodway and badway bends, respectively. The terms longitudinal and transverse for goodway and badway bending, respectively, are also used. Longitudinal and transverse refers to the direction of material movement, relative to the rolling direction, during bending.

Examples of bend formability in alloys C510 and C725 are provided by Figure 6. These curves of bend formability illustrate several characteristics of most copper alloys. The minimum bending radius increases (formability decreases) with increasing strength (or temper). Also, the badway (transverse) bend formability degrades more rapidly with increasing strength than goodway formability, as shown for C510. The comparison alloy, C725, is more

isotropic in formability. This comparison illustrates that directional uniformity in properties is both alloy and process dependent. The minimum bend radii shown can be affected by special processing to produce better formability in some alloys, such as in phosphor bronze.

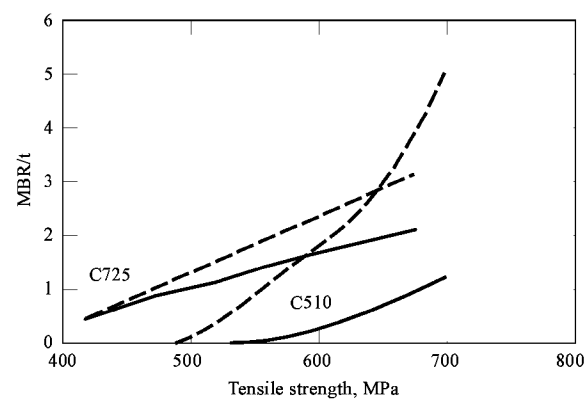


Fig. 6. The minimum radius over which alloy strip can be formed in bending without failure, normalized by thickness (MBR/ t) for alloys C510 and C725 as affected by increasing strength (temper) where (—) represents orientation of bending relative to the rolling direction of strip, ie, goodway (longitudinal), and (---) badway (transverse). To convert MPa to psi, multiply by 145.

Precipitation strengthened alloys show the same general trend, albeit at high strength levels. Typically, bend formability of these alloys is superior in the rolled temper condition, before aging treatment. Therefore, parts having the most demanding forming requirements can be first formed from rolled temper strip and then aged to obtain required strength.

Drawing. Copper alloys are often deep drawn or cupped. A second drawing step is commonly applied to the more severe or "deeper drawn" shapes required for some applications, such as ball point pen cartridge cases.

The deep drawability of alloy sheet is often determined from its capability of being drawn into a cylindrical cup (13–18). In this test, a circular disk or blank is held in place with a hold down ring over the female die of a punch and die set. Lubricant used between the hold down ring and the workpiece allows the disk to slide under the high loads generated as the blank is forced by the punch through the die. The test is run using blanks of various radii until the largest radius is found that does not fracture during drawing, which usually occurs at the base of the cup. The results of this test are expressed in terms of the formability parameter: the limiting drawability ratio (LDR), which is defined as the ratio of the largest circular blank diameter to the punch diameter. This parameter is used to compare the relative drawability of various alloys or various conditions of the same alloy.

This cupping test also provides a direct measure of anisotropy in the sheet resulting from crystallographic texture. Lobes, also called ears, that occur at the rim of formed cups are an indication of directional formability in the plane of the sheet. This effect is called planar anisotropy, the degree of which is measured from the average height of the ears. The location of earing with respect to the original rolling direction identifies the type of planar anisotropy. Earing is undesirable because it requires trimming and scrapping of metal from the rim area. Earing is directly related to the crystallographic texture and is reduced by control of texture through appropriate processing.

Plastic Strain Ratio. The plastic strain ratio is the ratio of strains measured in the width over the thickness directions in tensile tests. This ratio characterizes the ability of materials to resist thinning during forming operations (13). In particular, it is a measure of the ability of a sheet material to resist the thinning and failure at the base of a deep drawn cup. The plastic strain ratio is measured at 0°, 45°, and 90° relative to the rolling direction. These three plastic strain ratios R_0 , R_{45} , and R_{90} are combined to obtain the average strain ratio, called the $\bar{R}$ or the R value, and its variation in strain ratio, called ΔR :

$$\bar{R} = \frac{1}{4}(R_0 + 2 R_{45} + R_{90})$$
$$\Delta R = \frac{1}{2}(R_0 - 2 R_{45} + R_{90})$$

Much can be predicted from these parameters. Linear correlations have been established between the R value and the drawability parameter, LDR. Thus higher R values are associated with better deep drawability. ΔR , on the other hand, predicts the height and location of ears. For example, larger absolute values of ΔR predict higher ears during deep drawing cup shapes.

The sign of ΔR indicates the location of the individual ears with respect to the sheet rolling direction. For example, if ΔR is positive, the ears are located at 0° and 90° to the original rolling direction. This occurs when annealed copper has a pronounced cube crystallographic texture. When ΔR is negative, ears are at 45° to the rolling direction. This occurs when annealed fully recrystallized copper has a crystallographic texture that includes a significant component of the prior cold rolled texture. When annealed copper has no preferred crystallographic texture, called a random texture, ΔR is zero and there is no earing. But the random texture condition does not necessarily show the optimum R value, or drawability, that the material is capable of providing. The most desirable situation is for the ΔR parameter to be zero, predicting no plastic anisotropy, and the R value to be a maximum, predicting optimum drawability. In practice, it is difficult to achieve both situations with the same crystallographic texture. Over the years much work has been done with copper alloys and with brass in particular to develop mill processing schedules designed to provide sheet products having balanced textures that offer the best combination of high drawability and minimum earing.

Forming Limit Analysis. The ductility of sheet and strip can be predicted from an analysis that produces a forming limit diagram (FLD), which defines critical plastic strains at fracture over a range of forming conditions. The FLD encompasses the simpler, but limited measures of ductility represented by the percentage elongation from tensile tests and the minimum bend radius from bend tests.

The forming limit analysis involves deforming a clamped sheet specimen over a hemispherical punch into a dome shape until failure. At fracture, the dome height and the dimensional changes in a pre-gridded circle mesh pattern on the surface of a specimen are recorded, and the local strain pattern is calculated. From this strain pattern, the FLD is drawn, as shown in Figure 7 for C260 (14). Various deformation paths, ie, pure drawing (circles narrowed), pure bending (circles elongated), pure stretching (circles expanded), and in-between combinations modes, are produced by varying the test specimen width and the applied lubricant (13,14,17). The strains at the fracture, which are obtained from the distortion in the gridded circles closest to the fracture, are used to construct the forming limit diagram.

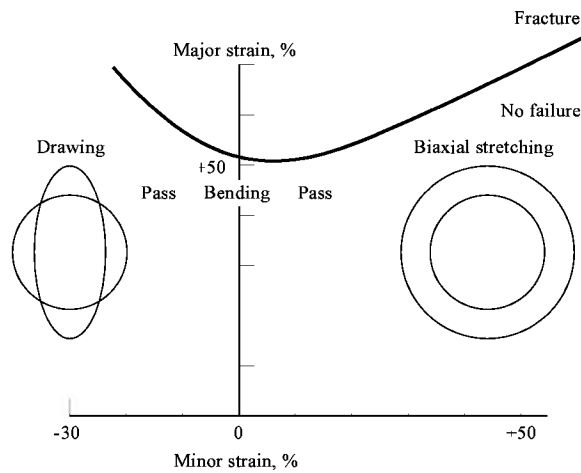


Fig. 7. The forming limit diagram (FLD) for C260-061 (annealed) showing limits to forming by various degrees of drawing (left of the origin), biaxial stretching (right of the origin), and pure bending (at the origin). Failure occurs at combinations of strains above the curve (14).

The region above the forming limit curve represents combinations of strain that lead to failure. Conversely, the region below the curve represents plastic deformation strains that the sheet material can withstand without failure. Drawing corresponds to a combination of positive and negative strain to the left of the diagram. Biaxial stretching (both strains are positive) is to the right of the diagram. The case of bending is in between with strain equal to zero in one direction.

A second diagram, the limiting dome height diagram (LDH), is also available from this analysis. The LDH diagram defines the limiting dome height obtainable for these various forming modes and as such is a measure of the overall ductility (14).

Springback. During most kinds of cold forming operations, regardless of material, elastic springback is encountered which can affect part dimensions (13). After metal is plastically deformed, elastic recovery occurs usually in the direction opposite to that of the plastic forming. The amount of springback can be predicted for simple bending, but prediction is more difficult for complicated shapes because plastic deformation is seldom homogeneous or uniform throughout formed parts.

The springback function, A_0/A_1 shown in Figure 8, is a guideline measure of springback. The angle to which the strip was initially formed by the tooling is A_0 , and may be 90° . The angle of bend that remains after forming appears as A_1 in this index and is 90° or greater. The value of 1.0 represents no springback. An increasing amount of springback is indicated by decreasing values of this parameter.

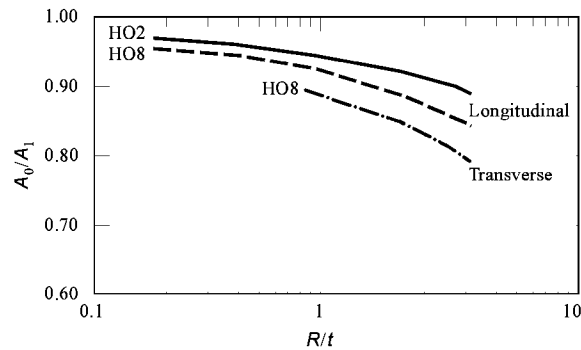


Fig. 8. Springback behavior for alloy C260, as defined by the ratio of the original forming angle, 90° or A_0 , divided by the final relaxed angle, A_1 . The effects of temper and orientation are shown as a function of bending radius R , normalized by the thickness of the strip t .

Manufacturers of copper alloy sheet and strip have developed springback data using this test as a guide for selecting alloys and tempers. The test is best applied to comparisons of springback among the various tempers of an alloy. By allowing for springback, forming tools and dies can be designed to ensure that dimensional specifications are met.

The degree of springback is higher for higher temper strip, for thinner strip, for larger radii of bending, and the direction of bending (longitudinal or transverse). As illustrated by Figure 8 for C260, the amount of springback increases with temper and increases with applied bending radius. The stronger tempers exhibit more directional behavior.

Softening Resistance. The ability of being readily annealed or softened in a controlled manner to restore ductility is beneficial in mill processing, but resistance to softening of the wrought product is often preferred during fabrication and subsequent service (19). Joining operations such as welding, brazing, and soldering are prime examples where softening resistance is essential. Most additions of alloying elements and impurities increase softening resistance. The amount by which softening resistance is increased is specific to the element being added and also depends on whether the element remains in solid solution or forms a second-phase particle.

Metallurgical mechanisms that impede the kinetics of thermally-induced changes, such as recrystallization, may also enhance softening resistance. The lowering of diffusion rates so as to delay the onset of recrystallization results from nickel and manganese additions. Segregation of solute to dislocations, sub-boundaries, and grain boundaries to impede motion and thereby prevent recrystallization results from phosphorus and zirconium additions. The presence of fine precipitates to pin dislocations and internal boundaries effectively interferes with recrystallization as in oxide-dispersed alloys. In general, finer and more homogeneous dispersions result in the most softening resistance.

In addition to compositional effects, softening resistance depends strongly on the amount of prior cold work; softening resistance is lowered with increasing amounts of cold work (Fig. 9). Thus the thermomechanical condition of the material must be known when applying softening data. Softening resistance data are best obtained using the pertinent service or process condition. Measurement of hardness or strength at room temperature after various exposure to times at temperature (isothermal data), or after exposure at various temperatures for a fixed time (isochronal data) provide the most useful, relevant softening information. For general comparisons between various alloys and conditions, the half-softening temperature is reported. This temperature is defined as that where hardness or strength (measured at room temperature) is lowered to half of its initial cold worked value after isochronal anneals (typically one hour). Half-softening temperatures of alloys shown in Figure 4, all having near 50% prior cold work are 220°C for alloy C110, 295°C for alloy C260, 340°C for alloy C510, and 525° for alloy C706.

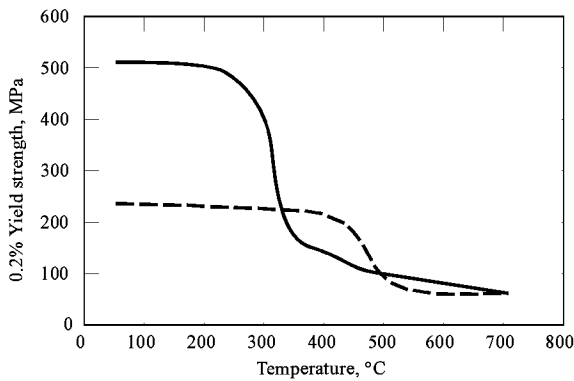


Fig. 9. The effect of prior cold work on softening resistance, as illustrated for C230 (copper–15% zinc) where (—) is CR 55%; (---) CR 10% in thickness. To convert MPa to psi, multiply by 145.

Stress Relaxation. Copper alloys are used extensively in applications where they are subjected to moderately elevated temperatures while under load. An important example is the spring member for contacts in electrical and electronic connectors. Critical to reliable performance is the maintenance of adequate contact force, or stability, while in service. Excessive decrease in this force to below a minimum threshold value because of losses in spring property can lead to premature open-circuit failure (see ELECTRICAL CONNECTORS).

Stress relaxation, while allied to creep, is different (20–23). Stress relaxation is the time-dependent decrease in load (stress) at the contact displacement resulting from connector mating. Here, the initial elastic strain in the spring contact is replaced by time-dependent microplastic flow. Creep, on the other hand, relates to time-dependent geometry change (strain or displacement) under fixed load, which is a condition that does not apply to connectors.

The rate at which elastic stress is reduced depends on the alloy, its temper, the temperature of exposure, and the duration. Resistance to stress relaxation of copper is improved by alloying with solid solution elements, as well as by dispersion and precipitation strengthening. Changing temper to higher strength for a given alloy results in some loss in stability. Relief annealing where yield strength is decreased slightly while causing little or no change in tensile strength is used to improve relaxation resistance.

Stress relaxation performance is usually determined in bending, where the initially imposed stress is 50 to 100% of the yield strength at the tension side of the stressed samples. Test details are presented in Reference 23. Data is presented as the percentage of the initial stress that remains as a function of time at exposure temperature.

Individual effects of alloying and processing on stress relaxation behavior are illustrated by Figure 10. Alloying with zinc or tin improves stress relaxation relative to copper; C260 (copper–30% zinc) and C521 (copper–8% tin) are solid solution hardened alloys. The phosphor bronzes offer higher resistance to stress relaxation than brass. Relief annealing also improves resistance to stress relaxation as illustrated by comparison of C521 in a rolled temper (H04) with the same material after relief annealing (HR04). Acceptable stability is often limited to 70% stress remaining or higher, depending on the application. Thus brass in the H04 temper would not be acceptable for applications at this temperature where expected duration of exposure can exceed 1000 hours.

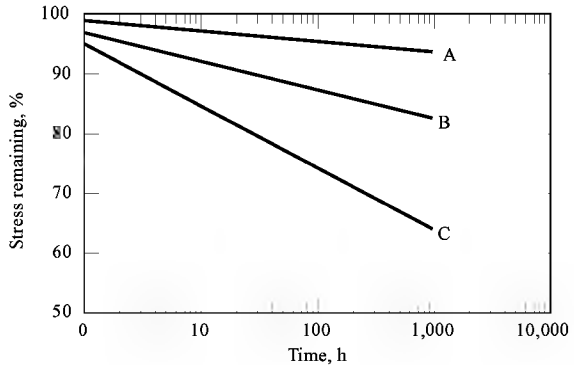


Fig. 10. Stress relaxation behavior at 105°C of the phosphor bronze alloy C521, in the A, relief annealed HR04 and rolled B, H04, tempers compared to C, rolled temper brass, C260-H04.

The highest stress relaxation resistance, at both high strength and moderate conductivity, is available from precipitation hardened alloys. This effect is an important consideration for applications that have high exposure temperatures, as found in present-day automobile engine compartments where 125–150°C exposure temperatures are possible. Precipitation hardened alloys C172 and C7025 are compared to a relief annealed solid solution hardened alloy, C521, in Figure 11.

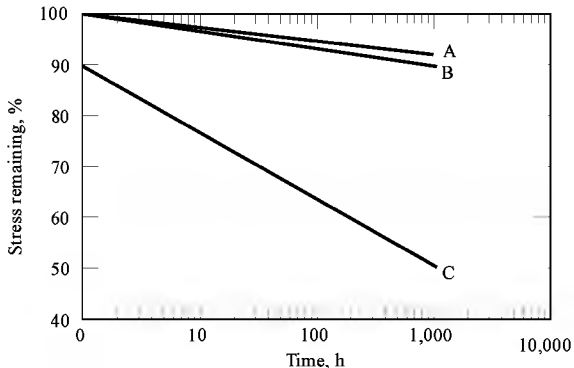


Fig. 11. Comparison between the stress relaxation resistance at 150°C, precipitation hardened alloys A, C172–TH02, and B, C7025–TM02, and C, a solution hardened alloy C521–HR02.

Fatigue. Imposed cyclic stressing of metals may result in localized cracking that leads to fracture. The pattern of stressing can be in reversed bending, reversed torsion, and tension–compression, or half cycles of these such as bending in only one direction. The number of cycles of stressing that can be endured without fracture depends on the magnitude of the peak applied stress, the pattern of stressing, and the alloy’s mechanical properties (20,24–28).

Fatigue properties are usually presented as graphs of applied stress, S , versus the logarithm of the number, N , of cycles to failure. The S – N curve for C510–H08 (spring) temper strip, tested in reversed bending, is shown in Figure 12 (25). Unlike steel, the S – N curve does not eventually become horizontal at high cycles so that an endurance limit does not exist. Thus fatigue resistance of copper and other face-centered cubic metals are reported as fatigue strength, which is defined as the applied stress at which failure occurs at 10^8 cycles.

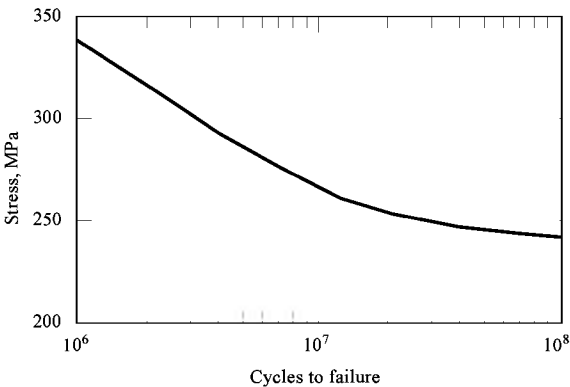


Fig. 12. Fatigue curve for C510–H08. To convert MPa to psi, multiply by 145.

Fatigue strengths for several copper alloys are listed in Table 7. Generally, fatigue strength increases with tensile strength of the material. This generality refers to the regime of high cycle fatigue where applied stress is a small fraction of its alloy's tensile strength (24). The rule-of-thumb is that the fatigue strength of a copper alloy is around one-third of its tensile strength.

Table 7. Fatigue Strengths of Copper Alloys^a

Alloy	Average 0.2% yield strength, MPa ^b	Fatigue strength 10 ⁸ cycles, MPa ^b
C172 ^c	760	275
C194	480	150
C260	615	185
C510	690	235
C762	725	205

^a All are H08, spring temper unless indicated.
^b To convert MPa to psi, multiply by 145.
^c TM02 (1 2 HM).

Fatigue properties in bending are most appropriate for copper alloys as these are often used as spring contact components in bellows and electrical switches and connectors. These articles are usually designed for acceptable service lives at a moderate to high number of stress cycles.

Leadframes used in electronic packages have a different requirement for resistance to fracture. These small exposed leads must be capable of sustaining only a few bending cycles without fracture but at high applied stress, should handling damage occur before the packages are assembled onto a circuit board. The resistance to low cycle fatigue by a material is controlled by ductility rather than strength (26). The ability to strengthen a damaged lead without its breaking is simulated in a lead bend fatigue test that is covered under Military Standard 883C (29). The test is done either on assembled packages or on appropriate samples prepared from alloy strip. Generally, lead materials are required to sustain at least three reversed 0–90° bends without fracture to ensure against scrapping packages.

Corrosion. Copper and selected copper alloys perform admirably in many hostile environments. Copper alloys with the appropriate corrosion resistance characteristics are recommended for atmospheric exposure (architectural and builder's hardware), for use in fresh water supply (plumbing lines and fittings), in marine applications (desalination equipment and biofouling avoidance), for industrial and chemical plant equipment (heat exchangers and condensers), and for electrical electronic applications (connectors and semiconductor package lead-frames) (30) (see PACKAGING).

Atmospheric exposure, fresh and salt waters, and many types of soil can cause uniform corrosion of copper alloys. The relative ranking of alloys for resistance to general corrosion depends strongly on environment and is relatively independent of temper. Atmospheric corrosion, the least damaging of the various forms of corrosion, is generally predictable from weight loss data obtained from exposure to various environments (31) (see CORROSION AND CORROSION CONTROL).

Pitting corrosion may occur generally over an entire alloy surface or be localized in a specific area. The latter is the more serious circumstance. Such attack occurs usually at surfaces on which incomplete protective films exist or at external surface contaminants such as dirt. Potentially serious types of corrosion that have clearly defined causes include stress–corrosion cracking, dealloying, and corrosion fatigue (27–34).

Stress corrosion is cracking that develops in sensitive alloys under tensile stress which is either internally imposed or is a residual after forming, in environments such as the presence of amines and moist ammonia. The crack path can be either intercrystalline or transcrystalline, depending on alloy and environment. Not all alloys are susceptible to stress corrosion (31).

The relative susceptibility of several commercial alloys is presented in Table 8. The index used is a relative rating based on integrating performance in various environments. These environments include the harsh condition of exposure to moist ammonia, light-to-moderate industrial atmospheres, marine atmosphere, and an accelerated test in Mattsson's solution. The latter testing is described in ASTM G30 and G37 (35,36) and is intended to simulate industrial atmospheres. The index is linear. A rating of 1000 relates to the most susceptible and zero designates immunity to stress corrosion.

Table 8. Relative Susceptibility to Stress Corrosion and Dealloying of Commercial Copper Alloys

UNS designation	Susceptibility	
	Stress corrosion ^a	Dealloying ^b
C260	1000	10

C230	200	4
C770	175	9
C422		4
C425	100	
C688	75	3
C510	20	
C521	10	
C706	0	
C194	0	0
C110	0	0

^a 1000 is most susceptible and 0 is essentially immune under normal service conditions.

^b 10 is most susceptible, 0 is least.

Dealloying refers to the selective removal of the more chemically active constituent of an alloy to leave a porous, weak deposit of the more noble constituent. Copper–zinc alloys containing more than 20% zinc are susceptible to dealloying, where it is called dezincification. Two-phase brasses having zinc contents near 39% are particularly prone to this form of attack. In brass, dezincification is easily apparent to the unassisted eye as a reddish copper "blush" on the surface of the metal (30,34).

The relative susceptibility to dealloying of alloys that contain varying amounts of zinc are also summarized in Table 8. The index of susceptibility was derived from testing in a 3.4% sodium chloride solution at 40°C. A ranking of 10 relates to high susceptibility, zero relates to immunity. The relative rankings among the alloys may differ for different exposure time and environment.

Corrosion fatigue refers to the combination of chemical attack and cyclic stressing, where cracking propagates more rapidly than under static stressing. Often, corrosion fatigue can be recognized by the presence of several cracks that initiate from stress raisers such as corrosion pits. The mode of fatigue is often transgranular.

Impingement or erosion attack can occur when liquids or gases impact metal surfaces at high velocity. The corrosion rate is high under such circumstances because any corrosion product films that can be protective if adherent are swept away as quickly as they are formed to leave exposed fresh surface.

Hydrogen Embrittlement. Copper alloys that contain cuprous oxide in their microstructures, as in C110, are potentially susceptible to embrittlement when heated in hydrogen-containing gases (37–39). The cuprous oxide is reduced by the hydrogen that readily diffuses into the alloy to produce water vapor that subsequently collects under substantial internal pressure, to cause internal pores. This porosity is typically found along grain boundaries, leading to fractures along these boundaries (39). Accordingly, susceptible alloys are annealed during processing in nitrogen or very low hydrogen-containing gas, at low temperatures, and short times, to avoid embrittlement.

Oxygen-free or deoxidized copper and its alloys are not susceptible to hydrogen embrittlement. Phosphorus is a commonly used deoxidizer (as in C122–phosphorus deoxidized copper). Aluminum, beryllium, magnesium, phosphorus, silicon, and zirconium can react with oxygen and cuprous oxide to form more stable oxides that are not readily reduced by hydrogen. These elements are often added to alloys to improve mechanical properties with resistance to hydrogen embrittlement being an added benefit. Examples of the latter are C151, C172, C194, C197, C260, C510, C688, and C7025.

Solderability. Most copper alloys have good solderability, meaning that they are wet by molten tin, tin–lead, and modifications of these to produce a continuous coating that has few to no pinhole sized nonwet areas. This characteristic of copper and its alloys accounts for the significant use of tin and solders to provide corrosion resistance and in joining (see SOLDERS AND BRAZING FILLER METALS) (40).

The inherent solderability of copper alloys for coating and joining purposes is determined by a variety of methods. The most commonly used is the vertical dip test where a sample is first fluxed, then immersed for a specific time (~5 s) in tin or solder, and finally visually inspected. This test procedure is provided in Military-Standard (41,42) and ASTM Specifications. A solderability rating system is summarized in Table 9. The flux that is used must also be specified. Fluxes used by the electronics industry are very mild and include resin (Type R) and mildly activated resin (Type RMA). Joining operations, as in radiator manufacture and plumbing, generally use much more aggressive chloride or acid-type fluxes because these applications are more tolerant to residual flux than is electronic equipment.

Table 9. Solderability Rating System and Alloy Ratings

Rating system class	Description	UNS designation ^a
I	uniform smooth coating, 100% wetting	C110, C194, C220, C510
II	>95% wetting	C195, C762
III	50–95% wetting	C260, C688
IV	<50% wetting	
V	no wetting	

^a Ratings are for a mildly activated resin (Type RMA) flux.

The inherent solderabilities of selected alloys are listed in Table 9. Class IV and V ratings with this particular flux indicate the presence of oxides or other surface contaminants that may be removable with more aggressive flux or acid pickling.

Tin and solder coatings (qv) are applied to copper and its alloys to provide the ability to be soldered after prolonged storage and to provide corrosion resistance (43–45). Such coatings are applied in several ways, each providing different coating characteristics. Coatings can be applied from a molten bath, wherein the strip surface is mechanically wiped as it exits the bath in order to produce a uniform coating thickness. Depending on the base alloy and process variables, the coating thickness is about 20–80 μm. Thicker coatings than those produced by mechanical wiping are made by directing strong air jets or "knives" at the surface. Reaction between the coating and the substrate copper alloy results in a 20–40 μm thick Cu–Sn intermetallic compound for both molten metal application methods. Finally, tin coatings can be applied by electrodeposition over a range of thicknesses of up to 300 μm. No intermetallic is formed during the plating process. However, rapid intermetallic growth subsequently occurs at room temperature, such that 20–30 μm thick intermetallic is present after 30 days. Further, electrotin coatings are often reflowed at just above the melting point by hot oil immersion or infrared radiation. Additional intermetallic compound forms during the reflow operation.

Coating characteristics affect performance. For instance, the intermetallic layer is easily oxidized during assembly and storage. If it is required that the material be soldered at some future time, it is necessary (but not always sufficient) for a residual layer of tin to be present. But this residual tin requirement is not necessary in many cases where the coating is intended to provide only tarnish resistance rather than extend shelf life solderability. Solder coatings are also degraded by diffusion of alloy constituents to the surface where they can react with the atmosphere to form sulfides, oxides, etc that impede subsequent soldering. The quality of tin coatings is evident by several accelerated tests, including a 150°C bake test and a steam aging test (42).

Brazing. Brazing is, by definition, elevated temperature soldering, that is, soldering above the arbitrarily defined temperature of 425°C. Copper and its alloys are readily brazed and often are brazed in order to take advantage of the stronger and more stable brazed joint compared to the soldered joint. In addition, it is easy to match the properties of the filler metal to the copper alloy to be brazed because many of the brazing alloys are themselves copper-base alloys (40). As with soldering, brazing is done at a temperature above the melting point of the brazing (filler) alloy and below that of the alloy being joined. The filler or brazing metal is placed in or near the joining surfaces in the form of a strip, wire, or powder. With application of heat, the filler

melts and by capillary action completely fills the gap and forms a metallurgical bond between the joining surfaces. For successful brazing it is necessary to set the proper clearance between the joining surfaces and to preclean the surfaces to be joined. In addition, oxidation protection must be provided via proper fluxing and annealing atmosphere to ensure a sound braze and adequate flow characteristics of the filler metal.

Filler alloys for brazing of copper alloys are usually copper base. These include copper–zinc alloys (RBCuZn), copper–phosphorus alloys (BCuP-1), and copper–silver (Zn, Cd, or Li) alloys (BAg). The copper–zinc brazing alloys are low in cost, but they require a flux and are not recommended for pure copper because of the potential for corrosion. The phosphorus-containing filler alloys are self-fluxing and are reasonably low cost, but are to be avoided with high nickel cupronickels because of an embrittling reaction between nickel and phosphorus. The copper–silver base alloys (BAg) produce good sound joints for most copper alloys, but carry the increased cost burden of the silver component.

The heating required for the brazing operation is readily done using commonly available industrial equipment: muffle furnace, oxy-gas torch, and induction heating. To protect against oxidation and to ensure good filler metal fluidity, fluxing agents and appropriate atmospheres are specified for each heating method. The fluxing agents comprise various mixtures of borates, fluorides, boric acid, and wetting agents. These agents are usually applied in paste form. The atmospheres used in brazing furnaces are usually exothermic (products of combustion), nitrogen, or dissociated ammonia, all having maintenance of low dew points. Whereas hydrogen can be used for many alloys, it must not be used for brazing electrolytic tough pitch copper, C11000, because of the embrittling steam reaction between hydrogen and the cuprous oxide particles. Moreover, the presence of hydrogen should be avoided during any brazing operation where cuprous oxide has the potential of forming.

Fluxes must be used when brazing the copper–zinc brasses. For copper alloys above about 20% zinc content, lower melting point brazes must be used to control zinc fuming and dezincification. Aluminum bronzes can be successfully brazed only when the proper flux (Type 4) is used to reduce the formation of alumina films. For the high nickel containing copper alloys it is best to avoid phosphorus-containing filler alloys to protect against lowered joint ductility resulting from formation of nickel phosphide. Various copper alloys can be successfully brazed only after they have received a stress relieving heat treatment and if they are heated slowly to the brazing temperature in order to avoid liquid metal embrittlement or fire cracking. Susceptibility to these latter problems are shown by the leaded alloys, the phosphor bronzes, the silicon bronzes, and the nickel silvers. When brazing precipitation hardening alloys, the brazing alloy and conditions can be chosen so that the alloy is solutionized during the brazing treatment. After a suitably rapid quench, the brazed alloy can be given an age hardening heat treatment.

Weldability. Welding has long been an important method of fabricating copper alloy parts as these alloys are generally readily weldable (40). But, as is the case with all materials, successful welding requires the proper match of welding technique to the specific application and copper alloy. There are three primary considerations for successful welding of copper. The first and foremost is the high thermal conductivity of copper and its resulting ability to conduct heat away from the weld zone. The second consideration is the chemical and in some cases toxic properties of the typical elements alloyed with copper. Adequate ventilation must be provided for the lead vapor, the zinc fumes, or the toxic compounds of Be, As, or Sb that can be emitted from the copper alloys containing these elements. The third concern is the ready solubility of oxygen in copper at elevated temperatures and the subsequent precipitation of cuprous oxide particles at grain boundaries during solidification and cooling, which if uncontrolled cause reduced strength and ductility in the weld zone.

Arc welding has long been used to join copper alloys. The gas tungsten-arc welding (GTAW) and the inert gas metal arc welding (GMAW) methods are the preferred arc welding methods for copper alloys. The GTAW technique is appropriate for workpiece thicknesses up to about 3 mm (0.125 in.) because of its capability of delivering intense heat into a narrow area. The GMAW technique is used for workpiece thicknesses greater than 12 mm (0.5 in.) because of its ability to provide rapid delivery of high heat. The rapid delivery of high heat is important for either technique to counter the high thermal conductivity of the high conductivity copper alloys. Besides heating efficiency, the importance of rapid delivery of the heat is to shorten the welding time to minimize oxygen pickup. The inert gases used are argon or mixtures of argon–helium. Although argon is cheaper, the helium content is often specified because of its higher thermal conductivity properties enabling higher heat input to the workpiece, especially for alloys with higher thermal conductivities.

In general, it is difficult to avoid porous welds of low ductility when arc welding oxygen-bearing electrolytic tough pitch copper, C11000. The situation is improved for arc welding of oxygen-free copper, but pickup of oxygen during welding must be avoided. Higher quality welds are more easily obtained when the workpiece and or the filler metal is a deoxidized alloy or otherwise contains deoxidizing elements. For this reason the phosphorus deoxidized alloys, the silicon and aluminum bronzes, and the cupronickels are arc weldable. Alloys having zinc contents above 20% fume too greatly; the free-machining alloys with Pb, Te, and S are prone to hot shortness (hot cracking) during cooling; alloys with Be, As, or Sb can emit toxic compounds.

In order to enable GTAW equipment to provide the required highly penetrating input of heat, a d-c straight polarity current is applied. Using such equipment, copper alloy tubes are welded to copper alloy tube sheets. The ends of processed strip or wire are easily joined in this technique to create longer coils of strip, wire, or rod product. When age hardening alloys are welded, such as the beryllium or chromium coppers, they must be given a post-welding age hardening heat treatment. Such a treatment usually does not return the weld region to the same strength as the rest of the workpiece, but it is adequate for most applications. Arc welding is not recommended for the newer oxygen dispersion hardened alloys (C15720), because the desirable uniform array of fine oxide particles (alumina) is destroyed by melting. Arc welding can be used for welding of copper to other metals, for which care must be taken to design for the differences in thermal expansion characteristics and during which the arc should be directed toward the higher conductivity alloy in the combination.

Resistance welding has been successfully applied to copper alloys in all of its various spot, seam, or butt joining modes. Because the process depends on ohmic (I^2R) heating at the interface to be joined, the ability to resistance weld is inversely related to electrical conductivity of the alloys being welded. The current is applied via opposing electrodes on either side of the joint. The electrodes themselves are usually made of high conductivity, softening resistant copper alloys, such as chromium– or zirconium–copper. The electrodes may be coated with a refractory metal, such as molybdenum, to prevent electrode pickup or sticking to the work piece. Most copper alloys having strip thicknesses on the order of 0.025–3.2 mm can be welded by this method. The leaded copper alloys are susceptible to liquid metal embrittlement as well as lead bleed out. As is the case for arc welding, steps must be taken to control the oxide fuming from alloys containing high levels of zinc, beryllium, and lead. Projection welding can be used to prevent electrode sticking and to control the shape of the weld nugget in high conductivity strip thicker than 0.5 mm. In this method, point, linear, or annular projections (ridges) are formed or machined in the workpiece to concentrate the welding current where the heat is needed. Flatter electrodes can then be used, leading to less electrode pickup.

Induction welding also depends on I^2R heating. In this case the electrical current is induced in the workpiece via an imposed high frequency magnetic field. The high frequency equipment is designed to concentrate the induced eddy currents near the surface at the edges of the workpiece to be joined. This method has been successfully applied to most copper alloys and, in particular, has been effectively used to make the seam in the high speed manufacture of welded copper alloy tube.

Machinability. Copper and its alloys can be machined with differing degrees of ease. Special additives such as lead, tellurium, selenium, and sulfur, are added, to enhance machinability (46), although other properties, such as formability and tensile ductility, normally suffer. These particular alloying elements form second-phase particles that promote chip fracture and the development of lubricative films at the tool-to-chip interface. Smaller, easier to handle chips and lower cutting forces having longer tool life result from use of these additives. Notable uses for special alloys having high machinability are rod, for high production rate screw machine items such as fasteners and plumbing components, and strip for keys.

Copper and its alloys are ranked according to a machinability rating index that is a percentage of the machinability of the most machinable copper alloy, C360, also known as free-cutting brass. The rating is based on relative tool life, where C360 is assigned a rating of 100. The actual performance of an alloy is dependent on the operation (turning, drilling, threading, etc), tool design, the lubricant, and the precise criterion used to define machinability (tool life, cutting rate, surface finish). Machinability ratings for several wrought alloys are given in Table 10.

Table 10. Machinability Ratings for Wrought Copper Alloys

UNS designation	Rating ^a	Comment
-----------------	---------------------	---------

C110	20	
C145	85	phosphorus deox, Te bearing
C147	85	sulfur bearing
C172	20	
C187	85	lead ed copper
C260	30	
C353	90	high lead ed brass
C360	100	free-cutting brass
C425	30	
C443	30	
C510	20	
C544	80	
C706	20	
C752	20	
C782	60	lead ed nickel silver

^a Rating is a relative scale based on C360 having a value of 100.

The characteristics of chips also follow the machinability rating. The most highly machinable alloys, ie, those having ratings near 60 and greater, develop short, brittle chips that are easily cleared from tools to make them suited to high production rate machining. Somewhat longer but brittle helical chips are often characteristic of alloys having ratings near 30, such as C260. The least machinable alloys, such as copper, phosphor bronzes, and the copper–nickel alloys having ratings of 20, form long, continuous chips that are easily entangled and difficult to handle.

Alloy Specific Properties

Copper. The physical properties of pure copper are given in Table 11. The mechanical properties of pure copper are essentially the same as those for C101 and C110. The coppers represent a series of alloys ranging from the commercially pure copper, C101, to the dispersion hardened alloy C157. The difference within this series is the specification of small additions of phosphorus, arsenic, cadmium, tellurium, sulfur, zirconium, as well as oxygen. To be classified as one of the coppers, the alloy must contain at least 99.3% copper.

Table 11. Physical Properties of Copper (C10100)^a

Parameter	Value
mp, °C	1083
density, g mL	8.94
electrical conductivity, % IACS min	101
electrical resistivity, μΩ·cm	1.71
thermal conductivity, W (m·K)	391
coefficient of thermal expansion from 20–300°C, μm (m·K)	17.7
specific heat, J (kg·K) ^b	385
elastic modulus, GPa ^c	
tension	115
shear	44

^a All properties at 20°C unless otherwise noted.

^b To convert J to cal, divide by 4.184.

^c To convert GPa to psi, multiply by 145,000.

Trace elements added to copper exert a significant influence on electrical conductivity. Effects on conductivity vary because of inherent differences in effective atomic size and valency. The decrease in conductivity produced by those elements appearing commonly in copper, at a fixed atomic concentration, rank as follows: Zn (least detrimental), Ag, Mg, Al, Ni, Si, Sn, P, Fe (most). Table 12 summarizes these effects. In the absence of chemical or physical interactions, the increase in electrical resistivity is linear with amounts of each element, and the effect of multiatom additions is additive.

Table 12. Solubility Limits and Electrical Conductivity Effects of Elemental Additions to Copper^a

Element	Solubility at room temperature, wt %	Resistivity increase, (μΩ·cm) wt%	Electrical conductivity ^b at 0.5 wt % ^a , % IACS
Ag	0.1	0.355	93
Al	9.4	2.22	62
As	6.5	5.67	38
Au	100.0	0.185	97
Be	0.2	4.57	44
Cd	0.5	0.172	98
Co	0.2	7.3	32
Cr	0.03	4.9	42
Fe	0.14	10.6	25
Mg	1.0	2.1	63
Mn	24.0	3.37	51
Ni	100.0	1.2	76
O	0.0002	21.0	14
P	0.5	14.3	20
Pb	0.02	1.02	79
Sb	2.0	2.9	55
Si	2.0	7.0	33
Sn	1.2	1.65	69
Ti	0.4	21.6	14
Zn	30	0.31	94
Zr	0.01	8.0	30

^a Refs. 9–12.
^b Assuming solubility at 0.5 wt %.

The mechanical properties of coppers having UNS designations between C10100 and C13000 are listed in Table 13. The coppers include high purity copper (C101, C102), electrolytic tough pitch (C110), phosphorus deoxidized (C120, C122), and silver-bearing copper (C115, C129, etc). The mechanical properties of these alloys are essentially the same. Other coppers, C142 through C157, offer higher strength at usually lower conductivity. The mechanical properties of alloys within the latter range are listed in Table 14.

Table 13. Tensile Properties of Copper (C10100–C13000)

Temper	Tensile strength, MPa ^a	Yield strength, ^b MPa ^a	Elongation, %	Hardness, HRF ^c
OS 025 ^d	235	76	45	45
H01	260	205	25	70
H02	290	250	14	84
H04	345	310	6	90
H08	380	345	4	94
H10	395	365	4	95

^a To convert MPa to psi, multiply by 145.
^b 0.5% offset.
^c Hardness is on the Rockwell-F scale.
^d Annealed to average grain size of 0.025 mm.

Table 14. Tensile Properties of Dilute Alloy Coppers (C14300–C15720)

Alloy	Temper	Tensile strength, MPa ^a	0.2% Yield strength, MPa ^a	Elongation, % 50 mm
C14300	H04	310	275	14
C14500 ^b				
C14700 ^b				
C18700 ^b	H04	330	304 ^c	20
C15000	TH04 ^d	445	411 ^c	18
C15100	H04	400	386	4
C15720 ^b	061	485	380	13

^a To convert MPa to psi, multiply by 145.
^b Rod of 25 mm diameter.
^c 0.5% offset for rod products.
^d Solution treated at 950°C, cold worked 90%, aged 1 h at 450°C.

Excellent resistance to saltwater corrosion and biofouling are notable attributes of copper and its dilute alloys. High resistance to atmospheric corrosion and stress corrosion cracking, combined with high conductivity, favor use in electrical electronic applications.

C101 and C102. The oxygen-free coppers, C10100 and C10200, are manufactured by melting prime quality, electrolytically produced cathode copper under reducing conditions designed to restrict oxygen content to below 0.0005% (C101) and 0.001% (C102). C101 has a minimum copper content, which may also include a trace of silver, of 99.99%. Oxygen-free copper is immune to hydrogen embrittlement, and provides low outgassing tendency, along with very high (101 % IACS min) conductivity at 20°C for annealed material.

C110. The most common commercial purity copper is C110. The principal difference between C110 and C102 is oxygen content which typically can be up to 0.05% in C110. Oxygen is present as cuprous oxide particles, which do not significantly affect strength and ductility, but C110 is susceptible to hydrogen embrittlement. The properties of C110 are adequate for most applications and this alloy is less costly than higher purity copper.

Deoxidized Copper. Cuprous oxide can be readily removed from copper by small additions of phosphorus, calcium, and magnesium. Phosphorus is the most common deoxidizing element. Most phosphorus-deoxidized coppers contain residual phosphorus in concentrations of 0.01–0.04%. The most common alloy in this family is C122, containing 0.015–0.040% phosphorus. Because of residual dissolved phosphorus, the minimum electrical conductivity of C122 is 85% IACS, versus 100 % IACS for C110. Applications involving brazing and welding in hydrogen-rich atmospheres are important uses for deoxidized copper where embrittlement by hydrogen has to be avoided.

Specialty Coppers. Additions are made to copper to satisfy specific needs. Tellurium at a nominal 0.5 wt % addition, sulfur at 0.35 wt %, and lead at 1 wt % enhance machinability. These alloys are identified as C145, C147, and C187, respectively. The solubility limit for each element is <0.001%, so that the excess is present as second-phase particles which assist in fracture of chips and lubrication during machining.

Additions of 0.1 wt % cadmium to phosphorus deoxidized copper (C143), or 0.1 wt % zirconium to copper (C151), significantly enhance softening resistance while maintaining very high (95 % IACS) conductivity. Because of environmental concerns, however, cadmium-containing alloys are being replaced by others that meet product and manufacturing requirements.

Extremely high softening resistance is available in alloy C15720 through dispersion strengthening by aluminum oxide. This alloy has been used successfully as a welding rod material and in other specific applications that can bear its cost.

High Copper Alloys. The high copper alloys contain a minimum of 96 wt % copper; most contain about 98.5 wt %. These alloys are found within the UNS classification of C162 to C197. As a group, they offer higher strength and better softening resistance than the specialty coppers. Like the coppers, the high copper alloys have excellent resistance to general and stress corrosion cracking. However, the high copper alloys have lower electrical conductivities than the coppers, they are more expensive, and special processing may be required to incorporate dispersoids or precipitates for optimal effectiveness. Nominal electrical conductivities and tensile properties of selected high coppers in hard rolled temper are summarized in Table 15.

Table 15. Conductivity and H04 (Hard) Temper Tensile Properties

UNS	Electrical conductivity, % IACS	Tensile strength, MPa ^a	0.2% Yield strength, MPa ^a	Elongation in 50 mm, %
<i>High copper alloys</i>				

C16200	90	414	310	5
C19000	60	579	531	3
C19200	65	448	414	7
C19400	65	462	434	4
C19500	50	593	572	2
C19700	80	448	414	7
<i>Cu–Zn</i> brass alloys				
C210	56	379	365	5
C220	44	428	400	4
C230	37	469	421	7
C240	32	496	421	4
C260	28	524	496	10
C353	26	503	462	12
C360 ^b	26	386	310	20

^a To convert MPa to psi, multiply by 145.

^b H02 Temper material.

Cadmium Copper (C162). Containing 1 wt % Cd, C162 has been used for many years in electrical connectors because of its excellent (90 % IACS) electrical conductivity and formability. Much of the cadmium is in solution; the remainder is dispersed as Cu₂Cd particles. Unique arc quenching ability, which reduces the surface degradation from electric-arc erosion during switching, is characteristic of this alloy. However, environmental concern because of cadmium's toxicity has all but eliminated its use, thus restricting its availability.

C190 to C197. The C190 to C197 alloys constitute a series of commercially useful copper alloys containing low levels of iron and phosphorus. They are relatively inexpensive and offer excellent combinations of strength, ductility, conductivity, and moderate softening resistance. As a group, these alloys are readily processed to provide a uniform array of iron and iron-phosphide dispersoids for hardening and grain size control. Offering electrical conductivities of 50–80 % IACS, they find extensive use in electronic products as leadframes and in electrical connectors. C194 (nominally, Cu–2.4 Fe–0.04 P–0.1 Zn) has the distinction of being the copper alloy that during the early 1970s replaced ASTM F30 (42 Ni–Fe alloy) as the dominant leadframe material in the electronics industry. Of increasing importance is the higher capacity for these alloys to dissipate heat from semiconductor packages and thus prolong integrated circuit (IC) life. Also of importance are their nonmagnetic properties permitting high clock speeds for IC devices (see INTEGRATED CIRCUITS; SEMICONDUCTOR TECHNOLOGY).

Copper–Zinc Brasses. Copper–zinc alloys have been the most widely used copper alloy during the 1990s. It is no accident that the word brass is included in the name of many copper alloy manufacturers. The manufacture of brass buttons and other brass artifacts was the principal reason for the establishment of the U.S. copper alloy industry in Connecticut during the 1800s.

Brass alloys fall within the designation C205 to C280 and cover the entire solid solution range of up to 35 wt % zinc in the Cu–Zn alloy system. Zinc, traditionally less expensive than copper, does not too greatly impair conductivity and ductility as it solution hardens copper.

Brass alloys are highly formable, either hot or cold, and provide moderate strength and conductivity. Moreover, the alloys have a pleasing yellow "brass" color at zinc levels above 20 wt %. The material is amenable to polishing, buffing, plating, and soldering. By far, the best known and most used composition is the 30 wt % zinc alloy, Cartridge Brass. Door knobs and bullet cartridges are the best known applications and illustrate the material's excellent formability and general utility. Properties of brass alloys are also summarized in Table 15.

The series of brasses, C312 to C385, contain from 0.25–5.0 wt % lead for the purpose of improving machinability. C360, having the composition Cu–35.4 Zn–3.1 Pb, has become the industry standard for machinability performance. It is given the descriptive name of free-machining brass and a machinability rating value of 100 (Table 10).

Brasses are susceptible to dezincification in aqueous solutions when they contain >15 wt% zinc. Stress corrosion cracking susceptibility is also significant above 15 wt % zinc. Over the years, other elements have been added to the Cu–Zn base alloys to improve corrosion resistance. For example, a small addition of arsenic or phosphorus helps prevent dezincification to make brasses more useful in tubing applications.

Tin Brasses. The tin brass series of alloys consists of various copper– (2.5 to 35 wt %) zinc alloys to which up to about 4 wt % tin has been added. These are solid solution alloys that have their own classification as the C40000 series of alloys. Tin provides better general corrosion resistance and strength without greatly reducing electrical conductivity. As with all the brasses, these alloys are strengthened by cold work and are available in a wide range of tempers. These alloys offer the combined strength, formability, and corrosion resistance required by their principal applications, namely, fuse clips, weather stripping, electrical connectors, heat exchanger tubing, and ferrules. Several tin brasses have lead additions to enhance machinability. For example, the high leaded Naval Brass C485, containing Cu–37 Zn–0.7 Sn–1.8 Pb, has a machinability rating of 78.

Table 16 illustrates the property enhancements and tradeoffs seen when tin is added to a copper–zinc brass base composition. The most commonly used alloys for electrical connectors are the Cu–10 Zn–Sn brasses, such as C411, C422, and C425. These lower level zinc–tin alloys offer good corrosion resistance along with the good formability, conductivity, and strength levels of brass.

Table 16. Conductivity and Wrought Tensile Strength of Tin–Brasses Showing the Hardening Effect of Tin Additions

UNS	Zinc, wt %	Tin, wt %	Electrical conductivity, % IACS	Tensile strength, MPa ^a	
				H04 Temper	H08 Temper
C411	10	0.5	32	448	552
C425	10	2.0	28	524	634
C220	10		44	428	524
C443	30	1.0	25	607	
C260	30		28	524	690

^a To convert MPa to psi, multiply by 145.

Admiralty Brass and Naval Brass are 30 and 40% zinc alloys, respectively, to which a 1% tin addition has been added. Resistance to dezincification of Cu–Zn alloys is increased by tin additions. Therefore, these alloys are important for their corrosion resistance in condenser tube applications. In these, as well as the other higher zinc compositions, it is common to use other alloying additives to enhance corrosion resistance. In particular, a small amount (0.02–0.10 wt %) of arsenic (C443), antimony (C444), or phosphorus (C445) is added to control dezincification. When any of these elements are used, the alloy is referred as being "inhibited." For good stress corrosion resistance, it is recommended that these alloys be used in the fully annealed condition or in the cold worked plus stress relief annealed condition.

Tin Bronzes. Tin bronzes may be the most familiar of copper alloys with roots going back into ancient times. Whereas bronze is still used for statuary, these alloys are found in many modern applications, such as electrical connectors, bearings, bellows, and diaphragms. The wrought tin bronzes are also called phosphor bronzes because 0.03 to 0.35 wt % phosphorus is commonly added for deoxidation and improved melt fluidity.

The several wrought alloys of commercial importance span the range of 1.0 to 10 wt % tin and are mostly used in work hardened tempers. These

alloys are single phase and offer excellent cold working and forming characteristics. Strength, corrosion resistance, and stress relaxation resistance increase with tin content. The tin–phosphorus alloys are highly resistance to stress corrosion cracking and have good resistance to impingement corrosion attack. Unfortunately, conductivity decreases, cost increases (tin has been historically more costly than copper), and the capability of being hot processed is impaired as tin level is increased. Only compositions up to around 4.5 wt % tin show good hot processing characteristics. The richer compositions are cast in thinner cross sections that can be cold rolled directly. Several cycles of working and annealing are employed to homogenize the cast microstructure.

One of the most highly used specialty tin bronzes is C544, containing Cu–4 Sn–4 Zn–4 Pb. Zinc provides increased strength to this tin bronze, whereas the lead addition provides good machinability (machinability rating of 80). The lead addition also enhances wear resistance. C544 is used in many applications where these two latter properties are of the utmost benefit, namely, bearings, bushings, gears, pinions, washers, and valve parts. Table 17 indicates the range of properties available for the commonly used phosphor bronzes.

Table 17. Conductivity and Wrought Tensile Strength of Tin Bronze Alloys

UNS	Tin, wt %	Electrical conductivity, % IACS	Tensile strength, MPa ^a	
			H04 Temper	H08 Temper
C505	1.0	48	441	552
C511	4.0	20	545	696
C510	5.0	15	572	738
C521	8.0	13	634	800
C524	10.0	11	690	882
C544	4.0 ^b	19	545	703

^a To convert MPa to psi, multiply by 145.
^b Also contain 4 wt % each of zinc and lead.

Aluminum Bronzes. Aluminum bronze alloys comprise a series of alloys (C606 to C644) based on the copper–aluminum (2–15 wt %) binary system, to which iron, nickel, and or manganese are added to increase strength. The mechanical properties of commercially important alloys in this group are listed in Table 18.

Table 18. Conductivity and H04 (Hard) Temper Tensile Strength of Aluminum Bronze Alloys

UNS	Constituents, wt %			Electrical conductivity, % IACS	Tensile strength, MPa ^a	0.2% Yield strength, MPa ^a	Elongation in 50 mm, %
	Al	Fe	Ni				
C614	7	2.5		14	586	400	35
C615	8		2	13	862	620	5
C625	13	4.5		10	690	379	1
C630 ^b	10	3.0	5	7	814	517	15

^a To convert MPa to psi, multiply by 145.
^b Also contains 1.5 wt % manganese.

Aluminum bronze alloys are used for their combined good strength, wear, and corrosion-resistance properties where high electrical and thermal conductivity are not required. Corrosion resistance results from the formation of an adherent aluminum oxide that protects the surface from further oxidation. Mechanical damage to the surface is readily healed by the redevelopment of this oxide. The aluminum bronzes are resistant to nonoxidizing mineral acids such as sulfuric or hydrochloric acids, but are not resistant to oxidizing acids such as nitric acid. However, these alloys must be properly heat treated to be resistant to dealloying and general corrosion.

The aluminum oxide surface layer that provides wear and corrosion resistance is not without drawbacks. This adherent film is difficult to remove during industrial cleaning of the alloys. Cleaning techniques have been developed that fragment the tightly adherent aluminum oxide film before removing it by the nonoxidizing mineral acids commonly used in brass mill processing. Furthermore, excessive tool wear in stamping and shearing equipment is caused by the presence of this film. Soldering and brazing are also made difficult by this oxide film.

Two single-phase, binary alloys are used commercially: C606, containing 5 wt % Al, and C610, 8 wt % Al. Both alloys have a golden color and are used in rod or wire applications, such as for bolts, pump parts, and shafts. Most of the available aluminum bronzes contain additional alloy elements, however. For example, C608 is a commonly specified condenser or heat exchanger tube alloy that is essentially 5% aluminum to which 0.02–0.35 wt % arsenic has been added to improve corrosion resistance. Alloy C613, having 7 Al–2.5 Fe also has a 0.3 wt % tin addition for improved resistance to stress corrosion.

Most of the aluminum bronzes contain substantial iron, nickel, or manganese additions. These alloying elements increase strength by forming second phases during heat treatment. Iron, the most commonly added element in these bronzes, separates as an iron-rich particle that controls grain size while it enhances strength. Nickel also reacts with aluminum to form NiAl precipitates during heat treatment with the same result as the iron addition.

Several Cu–Al–Si alloys have commercial importance. For example, C636 containing 3.5 Al–1.0 Si, is a rod and wire alloy that is processed to around 590 MPa 85,000 (psi) tensile strength. C638, containing 2.8 Al–1.8 Si–0.4 Co, is a strip product that owes its good strength and formability to dispersion strengthening and a very fine grain size that results from precipitation of cobalt silicide particles. The latter alloy attains tensile strengths of 566 MPa (82,000 psi) and 883 MPa (128,000 psi) in the annealed and spring rolled (H08) tempers, respectively.

Silicon Bronzes. Silicon bronzes have long been available for use in electrical connectors, heat exchanger tubes, and marine and pole line hardware because of their high solution hardened strength and resistance to general and stress corrosion. As a group, these alloys also have excellent hot and cold formability. Unlike the aluminum bronzes, the silicon bronzes have moderately good soldering and brazing characteristics. Their compositions are limited to below 4.0 wt % silicon because above this level, an extremely brittle phase (kappa) is developed that prevents cold processing. Electrical conductivities of silicon bronzes are low, however, because of the strong effect that silicon has in degrading this property. The three most popular of the commercially available alloys are listed in Table 19 with their typical properties.

Table 19. Conductivity and H04 (Hard) Temper Tensile Properties of Silicon Bronze Alloys

UNS	Si, wt %	Electrical conductivity, % IACS	Tensile strength, MPa ^a	0.2% Yield strength, MPa ^a	Elongation in 50 mm, %
C651 wire	1.5	12	690	483	11

C654 flat ^b	3.0	7	786	786	5
C655 flat	3.0	7	648	648	8

^a To convert MPa to psi, multiply by 145.

^b Also contains 1.5 wt % tin.

Modified Copper–Zinc Alloys. The series of copper–zinc base alloys identified as C664 to C698 have been modified by additions of manganese (manganese brasses and the manganese bronzes), aluminum, silicon, nickel, and cobalt. Each of the modifying additions provides some property improvement to the already workable, formable, and inexpensive Cu–Zn brass base alloy. Aluminum and silicon additions improve strength and corrosion resistance. Nickel and cobalt form aluminide precipitates for dispersion strengthening and grain size control. The high zinc-containing alloys are formulated to facilitate hot processing by transforming to a highly formable (beta) phase at elevated temperature.

Representative properties of alloys from this group are summarized in Table 20. C664, C688, and C690 are strip products that are used in electrical connectors, springs, and switches. These alloys are designed to provide beneficial combinations of strength and formability in cold rolled tempers. All three utilize dispersion hardening through iron plus cobalt particles (C664), cobalt aluminides (C688), and nickel aluminides (C690).

Table 20. Conductivity and Wrought Tensile Properties of Modified Copper–Zinc Alloys

UNS	Form	Composition, wt %	Electrical conductivity, % IACS	Tensile strength, MPa ^a	0.2% Yield strength, MPa ^a	Elongation in 50 mm, %
C664	flat	11.5 Zn–1.5 Fe–0.5 Co	30	607	586	4
C674	rod	36.5 Zn–2.8 Mn–1.2 Al–1.0 Si	23	634	379	20
C687	tube	20.5 Zn–2.0 Al–0.04 As	23	414	186	55
C688	flat	22.7 Zn–3.4 Al–0.4 Co	18	779	696	4
C690	flat	22.7 Zn–3.4 Al–0.6 Ni	18	772	703	5
C694	rod	14.5 Zn–4.0 Si	6	690	393	21

^a To convert MPa to psi, multiply by 145.

C674 and C694 are commonly used in rod and wire forms. These alloys are used in valve stems, shafts, bushings, and gears. Dispersoids of second phases that come about from beta-phase decomposition products account for their higher strength. C687, inhibited by the arsenic addition, is primarily a tube alloy that is used in condensers, evaporators, heat exchangers, and ferrules.

Copper–Nickels. The copper–nickel alloy system is essentially single phase across its entire range. Alloys made from this system are easily fabricated by casting, forming, and welding. They are noted for excellent tarnishing and corrosion resistance. Commercial copper alloys extend from 5 to 40 wt % nickel. Monel is a nickel–copper alloy that is outside of this range and contains 29–53 wt % of copper.

Iron is added in small (usually 0.5–1.0 wt %) amounts to increase strength. More importantly, iron additions also enhance corrosion resistance, especially when precautions are taken to retain the iron in solution. Precipitation of the iron–nickel-rich phase does not result in strengthening and can cause degradation of corrosion resistance (47). A small (up to 1.0 wt %) amount of manganese is usually added to both react with sulfur and deoxidize the melt. These copper alloys are most commonly applied where corrosion resistance is paramount, as in condenser tube or heat exchangers.

The common cupronickels are C704, which has 5 Ni–1.5 Fe; C706, 10 Ni–1.4 Fe; C710, 20 Ni; and C715, 30 Ni–0.7 Fe. Mechanical properties of the latter are summarized in Table 21. The best resistance to aqueous corrosion is available from C715 containing 30 wt % nickel. Both C715 and the lower cost C706 are suited for condenser and heat exchanger tubes in power plants. This group of alloys is also highly resistant to stress corrosion and impingement corrosion.

Table 21. Conductivity and H04 (Hard) Temper Tensile Properties of Cupro–Nickel Alloys

UNS	Alloy, wt %	Electrical conductivity, % IACS	Tensile strength, MPa ^a	0.2% Yield strength, MPa ^a	Elongation in 50 mm, %
C704	5.5 Ni–1.5 Fe	14.0	441	434	5
C706	10 Ni–1.4 Fe	9.0	531	517	2
C710	21 Ni	6.5	517	496	5
C715	30 Ni–0.5 Fe	4.6	565	538	3
C725	9 Ni–2 Sn	11.0	562	552	3

^a To convert MPa to psi, multiply by 145.

C725, a 9 wt % nickel alloy that is further strengthened by 2 wt % tin, is used in electrical connectors and bellows. Properties are summarized in Table 21. The alloy has good resistance to stress relaxation at room and moderately elevated temperatures, which accounts for its use in connectors and electrical circuit wire wrap pins.

Environmental factors must be considered when specifying copper–nickel alloys. The higher nickel compositions are less resistant to sulfides in seawater applications, for example. C713, containing 25 wt % nickel, is extensively used in U.S. coinage because of its good tarnish resistance. C722, containing Cu–17 Ni–0.8 Fe–0.5 Cr, is intended as a tube alloy having improved resistance to impingement corrosion in flowing water in condenser or heat exchanger applications.

Nickel–Silvers. Nickel–silver alloys, once called German silver, are a series of solid solution Cu–Ni–Zn alloys, the compositions of which encompass the ranges of 3 to 30 wt % Zn and 4 to 25 wt % Ni. This family of alloys falls within the UNS designation numbers C731 to C770. Leaded nickel–silvers that contain from 1.0 to 3.5 wt % lead for improved machining characteristics are designated as C782 to C799.

Nickel–silver alloys are not readily hot worked but have excellent cold fabricating characteristics. Because of the high nickel and zinc contents, these alloys exhibit good resistance to corrosion, good strength, and usually adequate formability, but have low electrical conductivity. Wire and strip are the dominant forms used for hardware, rivets, nameplates, hollowware, and optical parts. Table 22 lists common nickel–silvers and compares electrical conductivities and tensile strengths. Because of high nickel contents, these alloys are highly resistant to dezincification in spite of the high zinc content. Good resistance to corrosion in both fresh and seawater is another attribute.

Table 22. Conductivity and Hard Temper Tensile Properties of Nickel–Silver^a

UNS	Nickel, wt %	Zinc, wt %	Electrical conductivity, % IACS	Tensile strength, MPa ^b	0.2% Yield strength, MPa ^b	Elongation in 50 mm, %
C745	10	25	9	593	517	4
C752	18	17	6	586	565	5
C754	15	20	7	586	517	3
C757	12	23	8	586	517	4
C770	18	27	6	690	676	4

^a (Cu–Ni–Zn) alloys.
^b To convert MPa to psi, multiply by 145.

Of the several leaded nickel–silvers, C782 is one of the most commonly used. This alloy is not hot processible and has limited cold processing capacity. It is usually available as strip for application as keystock, watch parts, and plates because of its excellent corrosion resistance, hardness, and machinability.

Precipitation Hardening Alloys. Copper alloys that can be precipitation hardened to high strength are limited in number. In addition to the metallurgical requirement that the solubility of the added element(s) decrease with temperature, the precipitated phase that forms during aging must be distributed finely and have characteristics that act to resist plastic deformation.

Commercial precipitation hardening copper alloys are based on beryllium, chromium, and nickel, this last in combination with silicon or tin. The principal attributes of these alloys are high strength in association with adequate formability. Electrical conductivity varies according to alloy and ranges from around 20 to 80% IACS.

The precipitation hardening anneal can be done by the user of the material following fabrication of parts from solution treated–cold rolled temper strip. Because the precipitation hardening treatment decreases formability, it is possible to make the most geometrically demanding parts from rolled temper material, followed by aging to the highest hardness that the alloy can attain. Alternatively, the aging treatment can be done by the mill, to produce "mill hardened" strip, that represents a practical compromise between good formability and less-than-peak strength. Post-form cleaning operations are not necessary with mill hardened strip and formed tolerances can be closely maintained.

C170 to C172 (Beryllium Copper). The addition of beryllium to copper results in a significant age hardening response, making these alloys one of the few nonferrous materials that can reach 1400 MPa (200,000 psi) tensile strength (7). These alloys are more precisely ternary alloys. Cobalt or nickel serves as the third elemental addition. The purpose of this third element addition is to control grain size through the formation of coarse beryllides that pin grain boundaries and promote the hardening precipitation reaction.

Table 23 illustrates yield strength–bend formability trade-offs between a mill hardened temper (TM02) against aging rolled temper strip (H02) after forming to produce a TH02 temper product. Parts that require the same formability can be made using both the rolled temper and mill hardened material, but it is possible to achieve higher strength by aging after forming. However, aging after forming can introduce dimensional changes that accompany the hardening reaction. Moreover, the parts must be chemically cleaned with aggressive reagents to remove beryllium oxide formed during the thermal treatment. For these reasons, mill hardened strip represents the majority of the beryllium copper that is used.

Table 23. Properties of Beryllium Copper Alloys

UNS	Temper	Conductivity, % IACS ^a	Yield strength, MPa ^b	Formability, MBR t ^c	
				GW	BW
C172	H02	15	586	0.5	1
	TH02 ^d	20	1207	>5	>5
	TM02	18	758	0.5	1
C1751	TH04	50	758	2	2

^c For 90° bend forming angle.
^a Minimum values given.
^b To convert MPa to psi, multiply by 145.
^d Aged at 315°C for 2 h.

The trade-off between conductivity and formability, at comparable strength, is also shown in Table 23. The more dilute C1751 composition provides a higher conductivity in the mill hardened condition, but at a sacrifice in formability relative to mill hardened C172.

Toxicity concerns regarding beryllium result principally from possible consequences from inhaling its oxide. Any operation that produces airborne particles or dust of the oxide must be carried out using proper precautions to personnel.

C180 to C182 (Chromium–Copper). Chromium–copper alloys have widespread use in those applications, such as welding electrode tips, where good elevated temperature softening resistance is needed along with good electrical conductivity. As much as 0.65 wt % chromium can be dissolved in copper to produce a uniform array of pure chromium precipitates during the age hardening treatment. The low room temperature solubility of chromium (<0.03%) ensures good (usually about 85% IACS) conductivity after the subsequent age hardening heat treatment. Moreover, compositions above the maximum solubility level are commercially made in order to take advantage of the dispersion of coarser chromium particles that are useful for control of grain size. Typical properties of chromium–copper alloys in age hardened temper are summarized in Table 24.

Table 24. Conductivity and Age Hardened Tensile Properties of Chromium–Copper Alloys

UNS	Composition, wt %	Electrical conductivity, % IACS	Tensile strength, MPa ^a	0.2% Yield strength, MPa ^a	Elongation in 50 mm, %
C181	0.8 Cr–0.15 Zr–0.04 Mg	80	496	455	10
C18135	0.4 Cr–0.4 Cd	92	469	421	12
C182	0.9 Cr	80	462	407	14

^a To convert MPa to psi, multiply by 145.

C647 and C7025. Cu–Ni–Si alloys are precipitation hardened through the formation of Ni₂Si for optimum combinations of strength, conductivity, and formability. The composition and processing of C7025 also results in high temperature stability to both stress relaxation, needed for elevated temperature connector applications, and softening resistance, required for part manufacturing and service. These alloys are notable also for good general corrosion and stress corrosion resistance, that is equivalent to the highly resistant phosphor bronzes. High strength combined with moderate

conductivity available from the alloy are important for applications where ohmic (I^2R) heating from high electrical currents are possible.

Cu–Ni–Si alloys are also available in more dilute versions such as C647. Higher conductivity but lower strength are available from the latter alloy. Table 25 compares mechanical properties of the mill hardened condition of these alloys.

Table 25. Properties of Precipitation Hardened Cu–Ni–Base Alloys

UNS	Composition, wt %	Temper	Cond., % IACS	Tensile strength, MPa ^a	0.2% Yield strength, MPa ^a	Elongation in 50 mm, %
C180	2.5 Ni–0.6 Si–0.4 Cr	H50	48	690	525	12
C647	2 Ni–0.6 Si	TM04	52	615	520	6
C7025	3 Ni–0.65 Si–0.15 Mg	TM00	40	690	530	12
		TM03	35	785	735	7
C7265	7.5 Ni–5 Sn	TH04		1015	945	4
		TM04	15	860	755	12
C729	15 Ni–8 Sn	TH04		1240	1100	3
		TM04	8	860	790	15

^a To convert MPa to psi, multiply by 145.

C726 to C729. Very high strength, nearly equal to that available from beryllium–copper, is possible with precipitation hardened Cu–Ni–Sn alloys such as C729 (Cu–15 Ni–8 Sn). These alloys, referred to in the literature as being hardened through a type of precipitation reaction known as spinodal decomposition, encompasses a range of compositions that include Cu–9 Ni–6 Sn (C727) and Cu–7.5 Ni–5 Sn (C7265). More dilute compositions such as Cu–4 Ni–4 Sn (C726) are not precipitation hardenable. C727 and C729 are not easily hot worked and are usually made by strip casting followed directly by cold working, or alternatively by powder metallurgy processing.

Like beryllium copper, C729 is available in the rolled temper condition, for subsequent aging after forming, or in the mill hardened condition. Typical properties of these alloys are shown in Table 25. In addition, these alloys exhibit notably excellent resistance to stress relaxation at high application temperatures, for instance 200°C, and in this respect outperform beryllium–copper. However, the electrical conductivity of the strongest Cu–Ni–Sn composition (C729) is lower than that of C172.

Applications

Alloy selection for a particular application generally involves consideration of a combination of physical and mechanical properties as well as cost. Heat exchangers, including automobile radiators and domestic heating systems, are prime examples. Alloys used for such applications require not only good corrosion resistance and reasonable thermal conductivity, but also the capabilities of being formed into a variety of shapes and joined easily by soldering or brazing. C110 and C122 coppers are therefore widely used in such applications. C260 brass has also been used extensively in automobile radiators because of the good formability it possesses for manufacture of headers and because of its strength. Copper–nickel alloys such as C706 and C715 find application in condensers used in power utilities and desalination units because of their good corrosion resistance and reasonable conductivity relative to alternative materials.

Electrical interconnections that range from home wall receptacles to miniature connectors used in electronic products all require good formability and other properties specific to their application. Where the connector system is used in harsh environments, as an automobile engine compartment, resistance to stress relaxation is needed to provide stable contact force for reliable performance. Alloys like the beryllium coppers and C7025 are supplanting C260 brass and the phosphor bronzes in such elevated temperature applications. High conductivity alloys are favored in interconnection systems that carry large currents in order to avoid ohmic (I^2R) heating. Unacceptable loss in contact force and open-circuit failure of the connector could result if the operating temperature is too high for the alloy being used. The high copper alloys are therefore to be favored in such applications, provided that they also have adequate stress relaxation resistance. Alloys C151, C182, and C194 can more easily meet these property requirements.

Strength, thermal conductivity, formability, and the ability to be plated account for selection of particular alloys for use as leadframes in plastic encapsulated devices. Leads having high strength are needed to ensure against damage to the electronic package during handling before and during assembly to circuit boards. This requirement is particularly critical for high lead count packages where the leadframe material is thin (around 0.15 mm) and leads are extremely narrow (0.30 mm). The leadframe also provides a path for heat to be dissipated from the operating integrated circuit (chip) to prolong its service lifetime. Alloys that provide high strength near 690 MPa (100,000 psi) yield strength combined with high thermal conductivity (at least 175 W (m·K), equivalent to around 40% IACS electrical conductivity) are preferred for such applications.

An additional virtue of many copper alloys is the capacity for deep drawing such as cartridge cases and flexible bellows. C260 and C510 are well known for their drawability. Be–Cu that is drawn and then aged is used for bellows that have excellent fatigue life.

Finally, coinage and architectural uses take advantage of the variety of distinctive colors available from copper and its alloys, from red, through gold-yellow and silver-white, as well as tarnish and corrosion resistance. Among the alloys used are C713, a white-appearing alloy used in U.S. coinage; gold-appearing aluminum–bronze (C6155) and brass (C260); and red-brown copper–tin bronze. Clad coinage having dissimilar copper alloys for the clad and core components have the further advantage of being difficult to counterfeit when used in vending machines. The worldwide trend has been to use coins to replace low denomination paper bills because the useful lifetime of coins is much longer.

Economic Aspects

The relative pricing for selected high volume commercial alloys is illustrated in Figure 13, for which the price for C110 copper has been normalized to unity. The metal price of copper was \$0.52 /kg (\$1.14 /lb) at the end of 1992. Cost factors vary with changes in pricing of constituent alloy elements, order quantities, the supplier, and specified dimensional tolerances. Oxygen-free copper, C102, is sold as a premium product relative to copper. Brass, eg, C260, is less costly than copper because zinc is historically lower priced than copper. Tin-containing phosphor bronzes, eg, C510, are more costly than brass and copper because of higher tin metal cost, as are the copper–nickels, eg, C706, and nickel–silvers, eg, C752. Beryllium–copper alloys, such as C172, are the highest priced of the wrought copper alloys.

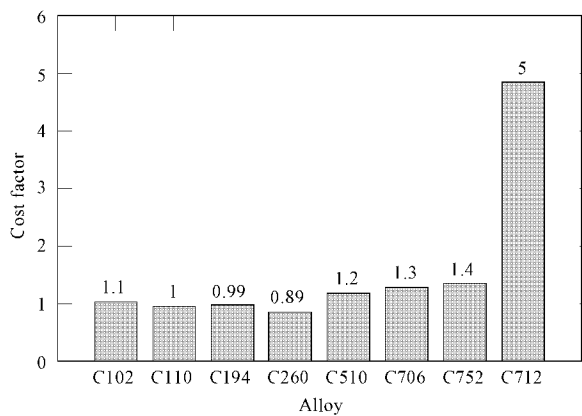


Fig. 13. Comparison of the costs of common copper alloys relative to that of copper. Actual cost factors vary according to individual alloy metal values, order size, and other factors.

BIBLIOGRAPHY

"Copper Alloys, Wrought" in *ECT* 1st ed., Vol. 4, pp. 431–458, by R. E. Ricksecker, Chase Brass & Copper Co.; "Copper Alloys (Wrought)" in *ECT* 2nd ed., Vol. 6, pp. 181–244, by R. E. Ricksecker, Chase Brass & Copper Co.; in *ECT* 3rd ed., Vol. 7, pp. 1–68, by R. E. Ricksecker, Chase Brass & Copper Co., Inc.

1. *Standard Practice for Temper Designations for Copper and Copper Alloys—Wrought and Cast*, ASTM B 601-92, American Society for Testing and Materials, Philadelphia, Pa., 1992.
2. "Alloy Phase Diagrams," Vol. 3, *Metals Handbook*, 10th ed., ASM International, Materials Park, Ohio, 1992.
3. T. B. Massalski, ed., *Binary Alloy Phase Diagrams*, Vol. 1, ASM International, Materials Park, Ohio, 1986.
4. M. M. Shea and N. S. Stoloff, *Met Trans.* **5**, 755 (1974).
5. R. L. Fleischer, *Acta Met.* **11**, 203 (1963).
6. *Standards Handbook, Part—Alloy Data Wrought Copper and Copper Alloy Mill Products*, 8th ed., Copper Development Association, Greenwich, Conn., 1985.
7. *The Mechanical Properties of Copper–Beryllium Alloy Strip*, ASTM STP 367, American Society for Testing and Materials, Philadelphia, Pa., 1964.
8. *Standard Test Method for Resistivity of Electrical Conductor Materials*, ASTM B 193-87, American Society for Testing and Materials, Philadelphia, Pa., 1992.
9. F. Pawlek and K. Reichel, *Zeit. Metallkunde* **47**, 347 (1956).
10. Y. T. Hsu and B. O'Reilly, *J. Met.* **29**, 21 (1977).
11. C. Drapier, *Revue de Metallurgie* **75**, 699 (1978).
12. J. W. Borrough, *Trans. Met. Soc. AIME* **221**, 1274 (1961).
13. *Forming and Forging*, Vol. 14, *Metals Handbook*, ASM International, Materials Park, Ohio, 1988, pp. 575–590, 809–824, 877–899.
14. E. Shapiro and F. N. Mandigo, *Forming Limit Analysis for Enhanced Fabrication*, INCRA Monograph X, *The Metallurgy of Copper*, INCRA, New York, 1983.
15. K. Suzuki, K. Ueda, and M. Tsuji, *Trans. Jpn. Inst. Met.* **25**, 716 (1984).
16. D. V. Wilson, *J. Inst. Met.* **94**, 84 (1966).
17. W. A. Backofen, *Met. Trans.* **4**, 2679 (1973).
18. K. Laue and H. Stenger, *Extrusion: Processes, Machinery, Tooling*, ASM, Materials Park, Ohio, 1976.
19. W. L. Finlay, *Silver-Bearing Copper*, Corinthian Edition, Copper Range Co., New York, 1968.
20. J. H. Mendenhall, *Understanding Copper Alloys*, Winchester Press, New York, NY, 1977.
21. A. Fox and E. O. Fuchs, *J. Testing Eval.* **6**, 211 (1978).
22. A. Fox, *J. Testing Eval.* **2**, 32 (1974).
23. *Standard Methods for Stress Relaxation Tests for Materials and Structures*, ASTM E 328-86, American Society for Testing and Materials, Philadelphia, Pa., 1992.
24. *Fatigue Crack Propagation*, ASTM STP 415, American Society for Testing and Materials, Philadelphia, Pa., 1967.
25. A. Fox, *J. Mater. JMLSA* **5**, 273 (1970).
26. C. E. Feltner and C. Laird, *Acta Met.* **15**, 1621 (1967).
27. J. P. Hickerson and R. W. Hertzberg, *Met. Trans.* **3**, 179 (1972).
28. P. O. Kettunen and U. F. Kocks, *Acta Met.* **20**, 95 (1972).
29. *Lead Integrity*, Method 2004, Dept. of Defense, Washington, D.C., 1977.
30. *Corrosion*, Vol. 13, *Metals Handbook*, 9th ed., ASM International, Materials Park, Ohio, 1987.
31. J. M. Popplewell and T. V. Gearing, *Corrosion* **31**, 279 (1975).
32. A. Sterling, A. Atrens, and I. O. Smith, *Br. Corros. J.* **25**, 271 (1990).
33. K. D. Eifird, *Corrosion* **33**, 3 (1977).
34. R. H. Heidersbach and E. D. Verink, *Corrosion* **28**, 397 (1972).
35. *Standard Test Methods for Use of Mattsson's Solution of pH 7.2 to Evaluate the Stress Corrosion Cracking Susceptibility of Copper–Zinc Alloys*, ASTM G 37-85, American Society for Testing and Materials, Philadelphia, Pa., 1992.
36. *Standard Practice for Making and Using U-Bend Stress—Corrosion Test Specimens*, ASTM G 30-84, American Society for Testing and Materials, Philadelphia, Pa., 1992.
37. S. Harper, V. A. Callcut, D. W. Townsend, and R. Eborall, *J. Inst. Metals* **90**, 423 (1961).
38. M. Koiwa, A. Yamanaka, A. Arita, and H. Numakura, *Mater. Trans., Jpn. Inst. Metals* **30**, 991 (1989).
39. *Standard Methods of Test for Hydrogen Embrittlement of Copper*, ASTM B 577, American Society for Testing and Materials, Philadelphia, Pa., 1992.
40. *Welding, Brazing, and Soldering*, Vol. 6, *Metals Handbook*, 9th ed., ASM International, Materials Park, Ohio, 1983.
41. *Wetting Balance Solderability*, Method 2022.2, Military Standard-883C, Dept. of Defense, Washington, D.C., 1987.
42. *Solderability*, Method 2003.5, Military Standard-883C, Test Methods and Procedures for Microelectronics, Dept. of Defense, Washington, D.C., 1987.
43. S. P. Zarlingo and J. C. Fister, "Solderability Performance of Tin Coated Copper Strip for Connector Components," *Proceedings, 41st Electronic Components and Technology Conference*, Atlanta, Ga., IEEE, 1991, p. 229.

44. *Specification for Electrodeposited Coatings of Tin*, ATM B 545-83, American Society for Testing and Materials, Philadelphia, Pa., 1992.

45. *Specification for Electrodeposited Coatings of Tin–Lead Alloy (Solder Plate)*, ASTM B 579-82, American Society for Testing and Materials, Philadelphia, Pa., 1992.

46. *Machining*, Vol. 16, *Metals Handbook*, 9th ed., ASM International, Materials Park, Ohio, 1989, pp. 805–819.

47. E. W. Palmer and F. H. Wilson, *J. Metal.* **4**, 55 (1952).

General References

Properties and Selection: Nonferrous Alloys and Pure Metals, Vol. 2, *Metals Handbook*, 10th ed., ASM International, Materials Park, Ohio, 1990, pp. 215–345.

Source Book on Copper and Copper Alloys, ASM International, Materials Park, Ohio, 1979.

Mechanical Testing, Vol. 8, *Metals Handbook*, 9th ed., ASM International, Materials Park, Ohio, 1985.

Powder Metallurgy, Vol. 7, *Metals Handbook*, 9th ed., ASM International, Materials Park, Ohio, 1984.

Copper '86 Copper Tomorrow, Conference Proceedings, Europa Metalli-LMI, Florence, Italy, 1986.

Copper '90 Refining, Fabrication, Markets, Conference Proceedings, Vasteras, Sweden, Institute of Metals, London, 1990.

Application Data Sheet—Standard Designations for Wrought and Cast Copper and Copper Alloys, Copper Development Association, Greenwich, Conn., 1992.

Copper and Copper-Base Powder Alloys, Metal Powder Industries Federation, Princeton, N.J., 1976.

E. A. Brandes and G. B. Brook, eds., *Smithells Metals Reference Book*, 7th ed., Butterworths, London, 1992.

Annual Book of ASTM Standards, Vol. 02.01, *Copper and Copper Alloys*, American Society for Testing and Materials, Philadelphia, Pa., 1992.

S. H. Butt and J. M. Popplewell, "Corrosion Considerations on the Selection of Materials in Automotive Terminal Systems," S.A.E. Paper 700031, Automotive Engineering Congress, Detroit, Mich., Jan. 1970.

K. Dies, *Kupfer und Kupferlegierungen in der Technik*, Springer-Verlag, Berlin, 1967.

G. R. Gohn, J. P. Guerard, and H. S. Freynik, *The Mechanical Properties of Wrought Phosphor Bronze Alloys*, ASTM STP 183, American Society for Testing and Materials, Philadelphia, Pa., 1956.

See also References 13, 17–20.

John F. Breedis
Ronald N. Caron
Olin Corporation

CAST COPPER ALLOYS

Copper [7740-50-3] (qv) alloy castings are used for their generally superior corrosion resistance (see CORROSION AND CORROSION CONTROL), high electrical and thermal conductivities, and good bearing and wear qualities. Some of the alloys are heat-treatable and couple high strength with good electrical and thermal conductivity. Irregular and complex external and internal shapes can be produced by various casting methods. The production of the same configurations by other methods may be mechanically impractical or too costly.

The cold- and hot-working properties of castings are not important because castings are usually not subject to mechanical forming operations; thus the chemical compositions of casting alloys have greater latitudes than those of comparable wrought alloys. However, the chemistry of the heat-treatable alloys must be carefully controlled to attain the proper hardening effects. Furthermore, compositions that do not lend themselves to forming operations are available in cast form for specific applications, such as in the bearing field for alloys containing up to 40% lead (see BEARING MATERIALS). Wrought methods for making the same product are unsatisfactory because the intermediate annealing of the alloy causes the lead to separate from the matrix and then from the material itself. The tolerance for impurities is normally greater in castings than in wrought alloys because the former are not mechanically formed.

However, in those cast alloys likely to be repaired or joined by welding (qv) some impurities can be very detrimental. On the basis of consumption, red brass alloys, Unified Numbering System (UNS) C 83600 (85% Cu, 5% Sn, 5% Pb, 5% Zn), UNS C 84400 (81% Cu, 3% Sn, 7% Pb, 9% Zn), and UNS C93200 (83% Cu, 7% Sn, 7% Pb, 3% Zn), are the most important of the cast copper alloys.

Properties and Characteristics

Cast copper alloys can be classified into two main groups: single-phase alloys, characterized by moderate strength, high ductility (except for leaded varieties), moderate hardness, and good impact strength; and polyphase alloys, having high strength, moderate ductility, and moderate impact strength.

Single-Phase Alloys. Copper–tin–zinc–lead alloys, tin bronzes, and the leaded tin bronzes have a narrow range of properties, namely, 234–303 MPa (34,000–44,000 psi) tensile strength, 103–152 MPa (15,000–22,000 psi) yield strength, and more than 20% elongation and reduction in area. Leaded alloys are lower in all the properties cited above. The fatigue strength of these alloys is, in general, in proportion to the tensile strength. Most of the alloys have a fatigue strength of about one-third times the tensile strength at 100 million cycles.

Alloys containing nickel and or aluminum have high strength and the impact data may exhibit scatter if the cooling rate or composition has not been closely controlled.

Polyphase Alloys. The two-phase alloys have a rather wide range of properties resulting from variations within the structure. If the second phase is distributed in critical depression, the hardness and strength are at a maximum and the ductility is at a moderate level. Tensile strength may be 415–825 MPa (60,000–120,000 psi); yield strength, 170–585 MPa (25,000–85,000 psi); and elongation, 10–40%.

The effect of a second phase is demonstrated in the copper–aluminum system, where increasing aluminum concentration causes the alloy system to change to a polyphase alloy. By obtaining a fine dispersion of the γ_2 phase, the yield strength is increased from 225 to 480 MPa (33,000–70,000 psi).

Alpha–beta aluminum alloys respond to heat treatment with a general improvement of mechanical properties. Heat treatment is accomplished by heating to 815–870°C, quenching in water, and reannealing at 370–535°C, depending on the size and section of the casting. Different combinations of strength, hardness, and ductility can be obtained. Some nickel in aluminum bronze is in solid solution with the matrix and helps refine the γ_2 precipitate, and a smaller amount is in the κ -intermetallic compound.

The desired balance of ductility and strength can be obtained in age-hardenable alloys, such as beryllium copper, by controlling the amount of precipitate. For higher strength, aging is conducted to provide a critical size dispersion. Greater amounts of precipitate are obtained by increasing the beryllium content of the alloy.

Copper–chromium and copper–nickel–silicon–chromium alloys are also precipitation hardenable. The precipitates are nickel silicides, chromium silicides, and elemental chromium. If conductivity is critical, the chromium–silicon ratio should be held at 10:1 so that appreciable amounts of either element are not left in solid solution in the copper after aging. Lithium can be used as a deoxidizer in copper alloys when conductivity is important. For a discussion of the principle of age- or precipitation-hardening copper alloys, see COPPER ALLOYS, WROUGHT COPPER ALLOYS.

Manganese bronzes form an interesting series in the alpha–beta area of the copper–zinc system. The mechanical properties are dependent on the relative amounts of the two phases and their distribution. Because the copper contents of the manganese–bronze alloys are relatively close together, they do not by themselves account for the presence of varying quantities of beta phase. The differences in beta content are accounted for by the effect of other elements on the location of the phase boundary between the alpha phase and the alpha–beta phase in the copper–zinc system. Aluminum, iron, and manganese all tend to shift the phase boundary from the value of 38% zinc. The shift of the phase boundary is given as $-0.87 \times \% \text{ Mn}$, $+1.0 \times \% \text{ Fe}$, and $+4 \times \% \text{ Al}$. The boundary shift for the manganese–bronze alloys may be computed and is shown in Table 1.

Table 1. Boundary Shift for Manganese–Bronze Alloys

Cu	Sn	Alloy composition ^a , wt %				Shift	New boundary, % Zn
		Pb ^b	Fe	Al	Mn		
58–63	0.5–1.5	0.8–1.5	0.5 ^c	0.5		0	38
58–62	0.5–1.5	0.5–1.3	0.8–1.5	0.25–1.0	0.1–0.5	1.98	36
55–60	1.0 ^c	0.3	0.4–2.0	0.5–1.5	1.5 ^c	2.80	35
60–68	0.1 ^c	0.1 ^c	2.0–0.4	3.0–7.5	2.5–5.0	22.32	16

^a Remainder is zinc in all cases.
^b Lead is insoluble and therefore does not enter into the calculation.
^c Maximum value.

This computation is also referred to as calculating the zinc equivalent of the alloy. The increase in strength in this alloy series is caused by increased amounts of beta phase in the structure. The silicon brasses show similar hardening effects accompanying a second phase. Typical mechanical properties and electrical conductivity for various cast alloys are shown in Table 2.

Table 2. Properties of Cast Copper Alloys^a

Common name	UNS designation	0.5% _o Yield strength, MPa ^a _b	Compressive yield strength, MPa ^a _b	Tensile strength, MPa ^a _b	Elongation in 5 cm, % _o	Brinell hardness, 500-kg load	Electrical conductivity, % IACS	Thermal conductivity at 20°C, W (m·K) ^c
copper	C80100	62		172	40	44	100	391
chromium–copper								
cast	C 81500	83		214	35	63	45	
precipitation-hardened		296		379	18	110	82	315
beryllium–copper	C 81700	469	551 ^d	620	8	217 ^e	48	188
ASTM B22								
high strength yellow brass	C 86300	124	413 ^f	820	83	225	8.0	35.4
gun metal	C 90500	138–158	276 ^g	310	25	75	11.0	74
tin–bronze 84:16	C 91100	172		214	2	135 ^e	8.5	
tin–bronze 81:19	C 91300	207		214		170 ^e	7.0	
high leaded tin–bronze	C 93700	124	90 ^f	214	20	60	10.0	47
ASTM B61								
steam–bronze	C 92200	137	262 ^d	2276	20	65	14	69
ASTM B62								
leaded red brass	C 83600	103	96 ^f	241	32	62	15	73
ASTM B66								
phosphor–bronze	C 94400	110	303 ^d	221	18	55	10	52
high leaded tin–bronze	C 93800	110	83 ^f	208	16	55	11.5	52
medium bronze	C 94500	83	248 ^d	172	12	50	10	52
high leaded tin–bronze	C 94300	90	76 ^f	186	10	48	9	62
ASTM B148								
aluminum–bronze 9A	C 95200	172–208	186–214 ^f	482–600	22–38	110–140	11	50
aluminum–bronze 9B								
cast	C 95300	208–241	110–138 ^f	482–586	20–35	110–160	13	62
heat-treated	C 95300	276–379	241–310 ^f	551–655	12–16	160–225 ^e		
aluminum–bronze 9D								
cast	C 95500	276–345	827 ^d	620–724	7–20	175–210 ^e	8.5	41
heat-treated	C 95500	413–551	1034 ^d	758–855	5–12	215–260 ^e		
aluminum–silicon–bronze	C 95600	234		517	18	140 ^e	8.5	38
manganese–aluminum–bronze	C 95700	310	1034 ^d	655	26	180 ^e	3.0	12
nickel–aluminum–bronze	C 95800	262	689 ^d	655	25	159 ^e	7.0	35
ASTM B176								
die-casting yellow brass	C 85800	207 ^h		379	15		20	
die-cast silicon–brass	C 87800	345 ^h		586	25		6.7	28
silicon–yellow brass	C 87900	241 ^h		482	25		15.0	
ASTM B584								
leaded red brass	C 83600	103	96	255	32	60	15.0	71.9
	C 83800	83–117	76–83	207–262	15–27	50–60	15.0	73
leaded semired brass	C 84400	90–117		200–269	18–30	50–60	18.0	73
	C 84800	103	90	234	30	55–59	16.4	73
leaded yellow brass	C 85200	83–96	55–96 ^f	241–276	25–40	40–50	15–22	83.9
commercial no. 1								
yellow brass	C 85400	76–103	62 ^f	208–262	20–35	40–60	18–25	88
yellow brass	C 85700	96–138		276–310	15–40	50–75	20–26	83.9
high strength yellow brass	C 86200	331	345 ^f	655	20	180 ^e	7.5	35.4
	C 86300	672 ^h	413 ^f	820	18	225 ^e	8.0	35.4
leaded high strength yellow brass	C 86400	172 ^h	158 ^f	448	20	105 ^e	19.0	88
high strength yellow brass	C 86500	193 ^h	165 ^f	489	30	130 ^e	22	86
leaded high strength yellow brass	C 86700	289		586	20	155 ^f	16.7	
silicon–bronze	C 87200	172	124 ^f	379	30	85	6.0	28.4
silicon–brass	C 87400	165		379	30	70	6.7	27.7
silicon–brass	C 87500	207	183 ^f	462	21	115	6.7	27.7
tin–bronze	C 90300	145	90 ^f	310	30	70	12	74.7
gun metal	C 90500	152	276 ^g	310	25	75	11	74.7
leaded tin–bronze	C 92300	138	241 ^d	276	25	70	12	74.7
high leaded tin bronze								
83:7-7-3	C 93200	117–148	317 ^d	207–262	12–20	67	12	58.1
85:5:9:1	C 93500	83–103	90 ^f	193–241	20–35	55–65	15	71
80:10:10	C 93700	124	122 ^f	241	20	60	10	46.9
78:7:15	C 93800	96–138	90–100 ^f	172–227	10–18	50–60	11.5	52
70:5:25	C 94300	76–103	83–96 ^f	158–207	7–16	72–55	9	62.6
nickel–tin bronze								
cast	C 94700	138–158		310–345	25–35	85	12	53.9
heat-treated	C 94700	345–413		517–551	5–10	180 ^e		
leaded nickel–tin bronze								

cast	C 94800	138–158	310–345	20–35	80	12	38.6
heat-treated		207	413	8	120 ^e		
	C 94900	103	262	15			
copper–nickel 90:10	C 96200	172	310	20		11	45
copper–nickel 70:30	C 96400	221	413	20	140 ^e	5	29
12° nickel–silver	C 97300	117	241	20	55	5.7	28.5
15° nickel–silver	C 97400	117	262	20	70	5.5	27.3
20° nickel–silver	C 97600	117	262	20	70	5	22
25° nickel–silver	C 97800	165	276	20	80	4.5	25.4

^a Mechanical property data were developed from separately cast test bars; values shown are based on technical literature.

^b To convert MPa to psi, multiply by 145.

^c To convert W (m·K) to Btu-ft² (ft·h·°F), multiply by 0.578.

^d 0.1 cm set cm.

^e 3000 kg load.

^f 0.001 cm set cm.

^g 0.01 cm set cm.

^h Offset of 0.2°.

Stresses and Stress Relieving. Nonuniform cooling leads to unbalanced residual stress patterns in the casting. Exposure of stressed castings to environments containing ammonia, ammoniacal compounds, or mercury can cause cracking of the alloys. Castings stressed to a level of ca 80 MPa (12,000 psi) and greater may crack in mercury environments, and stress levels of only a few MPa may produce cracks when exposed to ammonia. The coppers and copper–nickel–iron alloys usually have very high resistance to stress-corrosion cracking and do not require stress relieving.

Stress relieving to safe stress levels can be accomplished by a thermal treatment, sometimes called stress-relief annealing even though annealing may not be involved. The following alloys are all effectively stress relieved by heating at 260°C for 24 min cm of section: high copper alloys, red brass, semired brass, yellow brass, manganese–bronze, silicon–bronze, tin–bronze, and nickel–bronze. Aluminum–bronze alloys are treated at 316°C for 24 min cm of section.

If the casting is loaded or stressed externally in service, prior thermal stress treatment is of little value. Stress-corrosion cracking does not distinguish between residual or applied stresses.

Machinability. The cast copper alloys can be placed in three groups relative to machinability and are rated in the same general manner as wrought copper alloys. Group one contains the leaded alloys and is considered to be free-machining. Lead causes chip breakage during machining operations, permitting higher cutting speeds, decreased tool wear, and improved surface finish.

Alloys in the second group are polyphase alloys having a second phase generally harder than the matrix, which can cause brittleness and some chip breakage. This group comprises leaded tin bronze, silicon bronze, high tin bronze, aluminum bronze, and manganese bronze. Group three, the most difficult to machine, consists of high strength manganese and aluminum–bronze high in iron or nickel content. Machinability ratings are listed by alloy in Table 3.

Table 3. Machinability Ratings for Cast Copper Alloys^a

Common name	UNS designation	Machinability rating, ^b %
<i>Free-cutting alloys</i>		
leaded red brass	C 83600	90
	C 83800	90
leaded semired brass	C 84400	90
	C 84800	90
leaded yellow brass	C 85200	80
commercial no. 1 yellow brass	C 85400	80
yellow brass	C 85700	80
die-casting yellow brass	C 85800	80
high leaded tin–bronze	C 93200	70
	C 93500	80
bushing–bronze	C 93700	80
high leaded tin–bronze	C 93800	80
	C 94300	80
phosphor–bronze	C 94400	80
medium bronze	C 94500	80
12° nickel–silver	C 97300	70
<i>Moderately machinable alloys</i>		
leaded high strength yellow brass	C 86400	65
	C 86700	55
silicon–bronze	C 87200	40
silicon–brass	C 87400	50
silicon–brass	C 87500	50
die-cast silicon–brass	C 87800	40
stream–bronze	C 92200	40
leaded tin–bronze	C 92300	40
leaded nickel–tin–bronze		
cast	C 94800	50
heat-treated	C 94800	40
aluminum–bronze 9A	C 95200	50
aluminum–bronze 9B	C 95300	55
aluminum–bronze 9C	C 95400	60
aluminum–bronze 9D	C 95500	50
aluminum–silicon–bronze	C 95600	60
manganese–aluminum–bronze	C 95700	50

nickel–aluminum–bronze	C 95800	50
15° nickel–silver	C 97400	60
20° nickel–silver	C 97600	65
25° nickel–silver	C 97800	60
Difficult-to-machine alloys		
copper	C 80100	10
chrome–copper	C 81500	20
beryllium–copper	C 81700	30
high strength yellow brass	C 86200	30
	C 86300	8
	C 86500	25
	C 90300	30
tin bronze	C 90500	30
gun metal	C 91100	10
tin bronze 84:16	C 91300	10
tin bronze 81:19		
nickel–tin bronze		
cast	C 94700	30
heat-treated	C 94700	20
copper nickel 90:10	C 96200	10
copper nickel 70:30	C 96400	20

^a Ratings are compared with free-cutting brass—60° Cu, 3° Pb, 37° Zn—which has a rating of 100°.

Electrical and Thermal Conductivities. Electrical conductivity is customarily expressed as a percentage of The International Annealed Copper Standards (IACS) adopted in 1913 to represent the average electrical conductivity of high grade copper. Elements in solid solution with copper often have a marked effect on both the electrical and the thermal conductivity of the alloy. Alloying elements present in significant concentrations as well as low concentrations of deoxidized elements decrease both properties. Wrought alloys have higher conductivities than the comparable casting alloys; wrought alloys usually have lower concentrations and fewer alloying elements. Sound copper castings requiring at least an 85° electrical conductivity IACS require care in the preparation of their melts. Impurities such as iron, silicon, tin, zinc, aluminum, or a phosphorus deoxidizer cannot be used because small residual concentrations markedly lower conductivity. Calcium boride or lithium helps to produce sound castings having better conductivities.

The thermal conductivity of copper having an electrical conductivity of 100° IACS is 391 W (m·K) at 20°C. The Wiedemann-Franz ratio of thermal conductivity and the product of electrical conductivity times absolute temperature are approximately constant. Many copper alloys have increasing thermal conductivity with increase in temperature, whereas electrical conductivity decreases.

Electrical conductivity is comparatively easy to measure, whereas thermal conductivity is not. Electrical conductivity values for the important cast alloys are listed in Table 2. Figure 1 schematically shows the electrical conductivity of cast copper-base alloys compared with various other cast metals and alloys. The equation $Y = 4.184 + 3.93x$ gives an approximation of thermal conductivity in relation to electrical conductivity, where Y is in W (m·K) at 20°C and x is the ° IACS at 20°C.

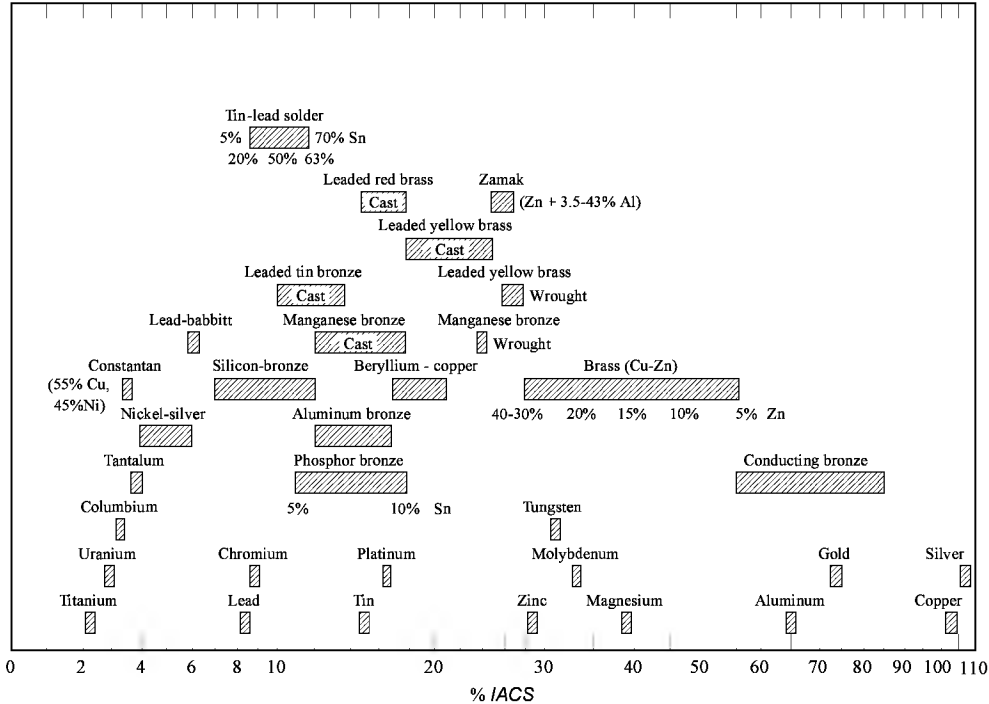


Fig. 1. Electrical conductivity of various metals and alloys. Courtesy of the American Foundrymen's Society.

Bearing and Wear Properties. Copper alloys have been used as bearing materials (qv) because of the combination of moderate to high strength, very good corrosion resistance (Table 4), wear resistance, and self-lubricating characteristics. Bearing alloys containing copper can be placed into three groups: phosphor-bronze alloys; copper-tin-lead alloys; and aluminum-bronze and silicon-bronze alloys.

Table 4. Corrosion Ratings of Copper Castings^a

	Co	Tin	Le	Hi	Le	Le	Lea	Le	Hi	Alu	Le	Le	Sili	Sili
	pp	bro	ade	gh	ade	ade	de	ade	gh	mi	ade	ade	co	co
	er	nze	d	lea	d	d	d	d	str	nu	d	d	n	n
			tin	de	red	se	yell	hig	en	m	nic	nic	bro	bra
			bro	d	bra	mir	ow	h	gth	bro	kel	kel	nze	ss
			nze	tin	ss	ed	bra	str	yell	nze	bra	bro		

	bronze				brass		ss	engineering brass	low brass		ss	size		
acetate solvents	B	A	A	A	A	A	B	A	A	A	A	A	A	B
acetic acid														
20° o	A	C	B	C	B	C	C	C	C	A	C	A	A	B
50° o	A	C	B	C	B	C	C	C	C	A	C	B	A	B
glacial	A	A	A	C	A	C	C	C	C	A	B	B	A	A
acetone	A	A	A	A	A	A	A	A	A	A	A	A	A	A
acetylene ^b	C	C	C	C	C	C	C	C	C	C	C	C	C	C
alcohols ^c	A	A	A	A	A	A	A	A	A	A	A	A	A	A
aluminum chloride	C	C	C	C	C	C	C	C	C	B	C	C	C	C
aluminum sulfate	B	B	B	B	B	C	C	C	C	A	C	C	A	A
ammonia, moist gas	C	C	C	C	C	C	C	C	C	C	C	C	C	C
ammonia, moisture-free	A	A	A	A	A	A	A	A	A	A	A	A	A	A
ammonium chloride	C	C	C	C	C	C	C	C	C	C	C	C	C	C
ammonium hydroxide	C	C	C	C	C	C	C	C	C	C	C	C	C	C
ammonium nitrate	C	C	C	C	C	C	C	C	C	C	C	C	C	C
ammonium sulfate	B	B	B	B	B	C	C	C	C	A	C	C	A	A
aniline and aniline dyes	C	C	C	C	C	C	C	C	C	B	C	C	C	C
asphalt	A	A	A	A	A	A	A	A	A	A	A	A	A	A
barium chloride	A	A	A	A	A	C	C	C	C	A	A	A	A	C
barium sulfide	C	C	C	C	C	C	C	C	B	C	C	C	C	C
beer ^c	A	A	B	B	B	C	C	C	A	A	C	A	A	B
beet-sugar syrup	A	A	B	B	B	A	A	A	B	A	A	A	B	B
benzine	A	A	A	A	A	A	A	A	A	A	A	A	A	A
benzol	A	A	A	A	A	A	A	A	A	A	A	A	A	A
boric acid	A	A	A	A	A	A	A	B	A	A	A	A	A	A
butane	A	A	A	A	A	A	A	A	A	A	A	A	A	A
calcium bisulfite	A	A	B	B	B	C	C	C	C	A	B	A	A	B
calcium chloride														
acid	B	B	B	B	B	B	C	C	C	A	C	C	A	C
alkaline	C	C	C	C	C	C	C	C	C	A	C	A	C	B
calcium hydroxide	C	C	C	C	C	C	C	C	C	B	C	C	C	C
calcium hypochlorite	C	C	B	B	B	C	C	C	C	B	C	C	C	C
cane-sugar syrup	A	A	B	A	B	A	A	A	A	A	A	A	A	B
carbonated beverages ^c	A	C	C	C	C	C	C	C	C	A	C	C	A	C
carbon dioxide														
dry	A	A	A	A	A	A	A	A	A	A	A	A	A	A
moist ^c	B	B	B	C	B	C	C	C	C	A	C	A	A	B
carbon tetrachloride														
dry	A	A	A	A	A	A	A	A	A	A	A	A	A	A
moist	B	B	B	B	B	B	B	B	B	B	B	A	A	A
chlorine														
dry	A	A	A	A	A	A	A	A	A	A	A	A	A	A
moist	C	C	B	B	B	C	C	C	C	C	C	C	C	C
chromic acid	C	C	C	C	C	C	C	C	C	C	C	C	C	C
citric acid	A	A	A	A	A	A	A	A	A	A	A	A	A	A
copper sulfate	B	A	A	A	A	C	C	C	C	B	B	B	A	A
cottonseed oil ^c	A	A	A	A	A	A	A	A	A	A	A	A	A	A
creosote	B	B	B	B	B	C	C	C	C	A	B	B	B	B
ethers	A	A	A	A	A	A	A	A	A	A	A	A	A	A
ethylene glycol	A	A	A	A	A	A	A	A	A	A	A	A	A	A
ferric chloride, sulfate	C	C	C	C	C	C	C	C	C	C	C	C	C	C
ferrous chloride, sulfate	C	C	C	C	C	C	C	C	C	C	C	C	C	C
formaldehyde	A	A	A	A	A	A	A	A	A	A	A	A	A	A
formic acid	A	A	A	A	A	B	B	B	B	A	B	B	B	C
freon	A	A	A	A	A	A	A	A	A	A	A	A	A	B
fuel oil	A	A	A	A	A	A	A	A	A	A	A	A	A	A
furfural	A	A	A	A	A	A	A	A	A	A	A	A	A	A
gasoline	A	A	A	A	A	A	A	A	A	A	A	A	A	A
gelatin	A	A	A	A	A	A	A	A	A	A	A	A	A	A
glucose	A	A	A	A	A	A	A	A	A	A	A	A	A	A
glue	A	A	A	A	A	A	A	A	A	A	A	A	A	A
glycerin	A	A	A	A	A	A	A	A	A	A	A	A	A	A
hydrochloric (muriatic)														
acid	C	C	C	C	C	C	C	C	C	B	C	C	C	C
hydrofluoric acid	B	B	B	B	B	B	B	B	B	A	B	B	B	B
hydrofluosilicic acid	B	B	B	B	B	C	C	C	C	B	C	C	B	C
hydrogen	A	A	A	A	A	A	A	A	A	A	A	A	A	A
hydrogen peroxide	C	C	C	C	C	C	C	C	C	C	C	C	C	C
hydrogen sulfide														
dry	C	C	C	C	C	C	C	C	C	B	C	C	B	C
moist	C	C	C	C	C	C	C	C	C	B	C	C	C	C
lacquers	A	A	A	A	A	A	A	A	A	A	A	A	A	A
lacquer thinners	A	A	A	A	A	A	A	A	A	A	A	A	A	A
lactic acid	A	A	A	A	A	C	C	C	C	A	C	C	A	C
linseed oil	A	A	A	A	A	A	A	A	A	A	A	A	A	A
liquors														
black	B	B	B	B	B	C	C	C	C	B	C	C	B	B
green	C	C	C	C	C	C	C	C	C	B	C	C	C	B

white	C	C	C	C	C	C	C	C	C	A	C	C	C	B
magnesium chloride	A	A	A	A	A	C	C	C	C	A	C	C	A	B
magnesium hydroxide	B	B	B	B	B	B	B	B	B	A	B	B	B	B
magnesium sulfate	A	A	A	A	B	C	C	C	C	A	C	B	A	B
mercury, mercury salts	C	C	C	C	C	C	C	C	C	C	C	C	C	C
milk ^c	A	A	A	A	A	A	A	A	A	A	A	A	A	A
molasses ^c	A	A	A	A	A	A	A	A	A	A	A	A	A	A
natural gas	A	A	A	A	A	A	A	A	A	A	A	A	A	A
nickel chloride	A	A	A	A	A	C	C	C	C	B	C	C	A	C
nickel sulfate	A	A	A	A	A	C	C	C	C	A	C	C	A	C
nitric acid	C	C	C	C	C	C	C	C	C	C	C	C	C	C
oleic acid	A	A	B	B	B	C	C	C	C	A	C	A	A	B
oxalic acid	A	A	B	B	B	C	C	C	C	A	C	A	A	B
phosphoric acid	A	A	A	A	A	C	C	C	C	A	C	A	A	A
picric acid	C	C	C	C	C	C	C	C	C	C	C	C	C	C
potassium chloride	A	A	A	A	A	C	C	C	C	A	C	C	A	C
potassium cyanide	C	C	C	C	C	C	C	C	C	C	C	C	C	C
potassium hydroxide	C	C	C	C	C	C	C	C	C	A	C	C	C	C
potassium sulfate	A	A	A	A	A	C	C	C	C	A	C	C	A	C
propane gas	A	A	A	A	A	A	A	A	A	A	A	A	A	A
seawater	A	A	A	A	A	C	C	C	C	A	C	C	B	B
soap solutions	A	A	A	A	B	C	C	C	C	A	C	C	A	C
sodium bicarbonate	A	A	A	A	A	A	A	A	A	A	A	A	A	B
sodium bisulfate	C	C	C	C	C	C	C	C	C	A	C	C	C	C
sodium carbonate	C	A	A	A	A	C	C	C	C	A	C	C	C	A
sodium chloride	A	A	A	A	A	B	C	C	C	A	C	C	A	C
sodium cyanide	C	C	C	C	C	C	C	C	C	B	C	C	C	C
sodium hydroxide	C	C	C	C	C	C	C	C	C	A	C	C	C	C
sodium hypochloride	C	C	C	C	C	C	C	C	C	C	C	C	C	C
sodium nitrate	B	B	B	B	B	B	B	B	B	A	B	B	A	A
sodium peroxide	B	B	B	B	B	B	B	B	B	B	B	B	B	B
sodium phosphate	A	A	A	A	A	A	A	A	A	A	A	A	A	A
sodium sulfate, silicate	A	A	B	B	B	B	C	C	C	A	C	C	A	B
sodium sulfide, thiosulfate	C	C	C	C	C	C	C	C	C	B	C	C	C	C
stearic acid	A	A	A	A	A	A	A	A	A	A	A	A	A	A
sulfur, solid	C	C	C	C	C	C	C	C	C	A	C	C	C	C
sulfur chloride	C	C	C	C	C	C	C	C	C	C	C	C	C	C
sulfur dioxide														
dry	A	A	A	A	A	A	A	A	A	A	A	A	A	A
moist	A	A	A	B	B	C	C	C	C	A	C	C	A	B
sulfur trioxide, dry	A	A	A	A	A	A	A	A	A	A	A	A	A	A
sulfuric acid														
78° ˆ or less	B	B	B	B	B	C	C	C	C	A	C	C	B	B
78 to 90° ˆ	C	C	C	C	C	C	C	C	C	B	C	C	C	C
90 to 95° ˆ	C	C	C	C	C	C	C	C	C	B	C	C	C	C
fuming	C	C	C	C	C	C	C	C	C	A	C	C	C	C
tanic acid	A	A	A	A	A	A	A	A	A	A	A	A	A	A
tartaric acid	B	A	A	A	A	A	A	A	A	A	A	A	A	A
toluene	B	B	A	A	A	B	B	B	B	B	B	B	B	A
trichlorethylened														
dry	A	A	A	A	A	A	A	A	A	A	A	A	A	A
moist	A	A	A	A	A	A	A	A	A	A	A	A	A	A
turpentine	A	A	A	A	A	A	A	A	A	A	A	A	A	A
varnish	A	A	A	A	A	A	A	A	A	A	A	A	A	A
vinegar	A	A	B	B	B	C	C	C	C	B	C	C	A	B
water														
acid mine	C	C	C	C	C	C	C	C	C	C	C	C	C	C
condensate	A	A	A	A	A	A	A	A	A	A	A	A	A	A
potable	A	A	A	A	A	A	B	B	B	A	A	A	A	A
whiskey ^c	A	A	C	C	C	C	C	C	C	A	C	C	A	C
zinc chloride	C	C	C	C	C	C	C	C	C	B	C	C	B	C
zinc sulfate	A	A	A	A	A	C	C	C	C	B	C	A	A	C

^a A is recommended; B, acceptable; C, not recommended (1).

^b Acetylene under pressure forms an explosive compound with copper when moist or when certain impurities are present. Alloys containing less than 65° ˆ Cu are satisfactory under this use. When gas is not under pressure other copper alloys are satisfactory.

^c Copper and copper alloys resist corrosion by most food products. Traces of copper may be dissolved and affect taste or color. In such cases, copper metals are often tin coated.

Phosphor–bronze alloys contain Cu, Sn or Cu, Sn, Pb, and have a residual phosphorus concentration of a few hundredths to 1° ˆ. Nickel can be added to refine the grain structure and is claimed to disperse the lead phase. Copper–tin bearings have high water resistance, high hardness, and moderately high strength.

An alloy containing 11° ˆ Sn, formerly referred to as gear bronze, is used, as its name implies, for making gears. Bronze alloys containing 18–20° ˆ Sn are successfully used for high loads and slow movement. A maximum load of 17 MPa (2500 psi) is permissible for the 17° ˆ Sn alloy. These alloys contain high amounts of phosphorus to improve hardness. Zinc in these alloys leads to seizing and galling and therefore is limited to a maximum of 0.25° ˆ.

Copper–tin–lead alloys are softer and are used as bearing material for lighter loads, below 5.5 MPa (800 psi), moving at moderate speeds. These materials include alloys UNS C 93700 (80° ˆ Cu, 10° ˆ Sn, 10° ˆ Pb), UNS C 93200 (83° ˆ Cu, 7° ˆ Sn, 7° ˆ Pb), and UNS C 93500 (85° ˆ Cu, 5° ˆ Sn, 9° ˆ Pb). Alloy C 93700 is an excellent general bearing material widely used in machine tools and in electrical and railroad equipment. Alloys C 93500 and C 93700

cost slightly less and are used in maintenance service. Alloys UNS C 93800 (78% Cu, 7% Sn, 15% Pb) and UNS C 94300 (70% Cu, 5.0% Sn, 25.0% Pb), containing higher quantities of lead, are used where high loads are encountered under poor operating conditions, that is, with poor or no lubrication, in a corrosive environment, or in dirty applications.

Aluminum bronze alloys containing 8–9% Al are widely used for bushings and bearings in light or high speed applications. Alloys containing 11% Al, both cast and heat-treated, are well suited for heavy service, such as in valve guides, rolling mill bearings, screw-down nuts, cams, and wear plates. As the aluminum content increases above 11%, the material becomes harder and stronger but ductility decreases. Alloys containing more than 13% Al have a Brinell hardness of 300 or greater, and the material is brittle. High (>13%) aluminum content alloys are often used as forming dies, particularly for forming stainless steel.

Manganese and silicon are frequently added to increase strength and hardness of aluminum bronze alloys. These elements form hard, intermetallic compounds that are imbedded in the softer matrix, producing a condition that enhances bearing and wear-resistance qualities. Aluminum bronzes generally have excellent bearing properties when used against steel. They operate efficiently with a minimum of lubrication and resist galling and scoring of the mating surface.

Joining Characteristics. Cast copper alloys may be joined by welding (qv), brazing, and soldering techniques with varying degrees of ease and success (see also SOLDIERS AND BRAZING ALLOYS). Table 5 lists a number of copper-base casting alloys and indicates the ease with which they can be joined by various methods.

Table 5. Joining Characteristics of Cast Copper Alloys^a

Common name	UNS designation	Weldability			Braz-abil ty	Solder-abil ity
		Oxyacet-ylen e	Carbon arc	Metal arc		
copper	C 80100	NR	F	NR	E	E
chromium–copper	C 81500	NR	F	NR	G	G
beryllium–copper	C 81700	NR	NR	NR	G	G
leaded red brass	C 83600	NF	NR	F	G ^b	E
leaded semired brass	C 84800	NR	NR	F	G ^b	E
steam bronze	C 92200	NR	NR	NR	E ^b	E
tin bronze						
leaded	C 92300	NR	NR	NR	G ^b	E
high leaded	C 93200, C 93500, C 93700	NR	NR	NR	G ^b	G
	C 93800, C 94300, C 94500	NR	NR	NR	P ^b	G
phosphor bronze	C 94400	NR	NR	NR	G ^b	G
high strength yellow brass	C 86200	G	NR	G	P	P
	C 86300	P	P	G	P	P
	C 86500	P	P	P	F	F
leaded aluminum bronze	C 86400	P	P	P	F	F
9B ^c	C 95300	NR	F	G	G	G
9C ^c	C 95400	NR	G	E	E	G
9D ^c	C 95500	NR	P	G	F	G
silicon bronze	C 87200	G	P	F	F	NR
silicon brass	C 87500	F	NR	NR	F	NR
nickel silver	C 97300, C 97400, C 97600, C	NR	NR	NR	E	E
	97800					

^a E, excellent; G, good; F, fair; P, poor; NR, not recommended.

^b Strain must be avoided during brazing and cooling to control cracking.

^c Aluminum bronzes must use special flux for brazing and soldering.

Mechanical Properties. Most alloys containing tin, lead, or zinc have moderate tensile and yield strengths and high elongation. Higher tensile or yield strengths are available through the use of aluminum and manganese bronzes, silicon brasses and bronzes, and some nickel alloys. Some of these alloys, such as beryllium–copper, chromium–copper, and some aluminum–bronze, are heat-treatable to attain maximum tensile strengths. Mechanical and physical properties for copper-base casting alloys are summarized in Table 2.

The entire group of leaded alloys are the easiest to cut and machine. Most cast alloys can be joined by welding, brazing, and soldering (Table 5).

Copper–tin bearings have high resistance to wear, high hardness, and moderately high strength. Copper–tin–lead alloys are used when a softer material is needed and moving parts are at moderate speeds and loads are light. High strength manganese–bronze alloys have high tensile strength, hardness, and resistance to shock. General uses are for slow-motion and relatively high compression applications.

Production Methods

Precautions. Care should be taken not to impair the quality of the metal as a result of the melting operation. The best practice is to select scrap of known good quality. The use of miscellaneous scrap may not be economical, despite the lower cost of such material; however, it is not common practice to make much use of virgin metals.

Ingots produced by secondary smelters and refiners and made to specifications are a good source of melting stock. Scrap, such as sprues and gates, and turnings from the foundry's own castings are very acceptable melting materials.

The melt may be exposed to gaseous and solid impurities unless suitable precautions are taken. The most deleterious of these contaminants are hydrogen or water vapor, which may produce a porous structure in the cast alloy and ruin its quality. Oily scrap may also cause gaseous contamination of the melt. If an open flame can come in contact with the melt, care should be taken to ensure a slightly oxidizing flame, preferably containing 0.1–0.2% excess oxygen.

Electric furnaces do not utilize gas as fuel, and it is possible to operate with a near-neutral atmosphere. Care must be taken not to introduce water with the charge. Damp fluxes, charcoal, and other melt additives are sources of water contamination; patched and newly lined furnaces or ladles must be thoroughly dried before use. Some foundries use electrical strip heaters to slowly remove most of the moisture and use torch flames to complete drying and fusing the surface of the refractory. Rapid drying may lead to cracked refractory. Most molding methods can be used for the production of copper alloys. The choice of molding methods is based on cost, surface finish requirements, and tolerances.

Sand Casting. Sand casting is the most popular method of molding for cast copper alloys. It is the least expensive near net shape process, and pattern costs are low. Dimensional tolerance capabilities vary widely and are the least accurate when compared with other methods of casting.

Sand (usually silica), clay (bentonite), and water are the basic mixture of ingredients for green sand casting. The clay coats the sand and bonds the mixture. Water is used to help coat the clay and activate its binding capabilities. Binders for the sand are generally two-part resin–catalyst systems. The catalysts can be in liquid or gaseous form.

The dimensional capabilities and tolerances of chemically bound systems are better than those of green sand. The cost for producing components with these methods is 1.5 to 2 times higher than the cost of producing green sand.

Permanent-Mold Castings. The permanent mold is a reusable die usually made of cast iron, high alloy steel, or beryllium–copper. The process utilizes both gravity and low pressure pouring techniques and can be semiautomated. Sand cores can be used for internal coring. The molds are expensive to produce; therefore, the process is used when large quantities of castings are required. Some of the advantages over sand casting are tighter dimensional tolerances, excellent cast surface finish, higher mechanical properties, better internal soundness, reduced machining costs, and higher yields. Permanent molding is usually limited to short freezing range alloys such as aluminum, silicon, and manganese bronzes. Yellow brasses are also commonly used. Generally, tolerance control is within 0.010 cm cm, although closer control is possible.

Die Casting. Die casting is a process by which molten metal is injected into a metal die under pressure. Automatic or semiautomatic cyclic operations are normally employed. The die assembly consists of two halves with metal cores. The die-casting machine consists of a press, a special melting furnace, an injection mechanism, and necessary controls. The die halves are fixed to the opposing platens of the press so that they match when the press is closed. The other moving parts, including the injection mechanism designed to introduce molten metal at high velocity and under pressure into the die cavity, are integrated with the press so as to operate automatically in proper sequence as the press opens and closes.

The proper coordination and regulation of such factors as speed and pressure of the injection plunger, temperatures of both the dies and the metal, and gating and venting influence the quality of die castings, especially with regard to soundness of structure and surface finish.

Die casting is a special form of permanent-mold casting. Cost is an important factor in selecting this process. Yellow brass (UNS C 85700) is the largest volume alloy to be die cast, although other permanent-mold alloys may be used. The method is used when a difficult cored section or relatively large surface area to volume components are required. Cast sections having tolerances of 0.070 cm cm of thickness are not difficult to obtain. Die casting is best suited for large numbers of small, intricate pieces.

Plaster-Mold Casting. Plaster-mold casting is similar to sand casting but uses gypsum plaster as the molding medium. It uses a match plate or cope and drag patterns, cope and drag mold assemblies, cores, and gravity pouring. The tooling is relatively low in cost. The plaster process provides casting of superior surface finish, accurate reproduction, high dimensional accuracy, and reduced machining costs. High volume production is limited to small-parts applications that can be automated. The process is used extensively for prototyping in advance of die and permanent-mold casting designs. Nonleaded copper alloys, such as aluminum bronze, yellow brass, manganese bronze, silicon bronze, and low nickel bronze, are best suited for plaster-mold casting. Lead in the alloy must be kept to a minimum because the lead reacts with the plaster thereby causing defects. Tolerances of within 0.005 cm cm can be achieved. Parting line tolerances are within 0.10 cm cm.

Centrifugal Casting. Centrifugal casting consists of pouring the melt into a rotating mold, which may be in either a horizontal or vertical position. Most alloys can be cast by this method. Size of castings is not a constraint. Castings from a hundred grams to several metric tons may be produced. Clean castings can be produced because dross and nonmetallics, which are lighter than the base metal, are thrown toward the surface, where they can be removed by subsequent machining. However, machining and scrap cost must be considered. Soundness is enhanced by high pressures and consequent effective feeding during the solidification process.

Mechanical properties of centrifugally cast copper alloys vary with the alloy and mold material. Centrifugal castings made in sand molds, where relatively slow cooling prevails, have poorer mechanical properties than castings made in metal or graphite molds. However, even centrifuged castings made in sand molds have higher tensile strength than the same alloy and configuration cast statically in conventional sand molds. The combination of handling molten metal and rotating equipment emphasizes the need for exercising safe operating practice and caution in the casting process.

Investment Casting. Investment casting, also known as the lost-wax process, is a method for making intricate or complex castings to close dimensional tolerances. The process consists of making a pattern from materials such as plastic or wax. The patterns are assembled onto a cluster, dipped in a ceramic slurry, stuccoed with aggregate, then allowed to dry. Slurry dip and stucco steps are repeated until a 0.64–0.96 cm shell has been built. The wax is removed by flash firing or autoclaving. The shell is then fired in an oven to develop strength. Metal is melted and poured into a hot shell.

The investment-casting process offers extreme precision, excellent surface finishes, and part-to-part consistency. This method is the most costly method of producing cast components to near net shape. Its value is realized through the high production of complex castings.

Continuous Casting. Continuous-casting methods are being applied to the production of copper-base alloys. The molten metal is continuously poured into the top of the water-cooled, lubricated mold, and the solid cast shape is continuously withdrawn mechanically from the bottom of the mold. The process is continuous as long as molten metal is available and the mold does not wear out. It may be made semicontinuous by casting a few meters in length, then stopping the process and removing the casting from the system; after suitable mold preparation, the process can be restarted.

Where more than one strand of casting is required, a multihole, water-cooled die can be fixed to the bottom of a holding furnace and the metal withdrawn as a solid strand from the bottom of the furnace, either in a horizontal or vertical plane.

The process requires considerable equipment, and skills in casting must be acquired. Continuous casting is useful where soundness and high volume of parts are needed. The process is also used to produce shapes, such as billets, bars, and cakes, as starting materials for the wrought-product industry.

Comparison of Casting Methods. Table 6 compares several factors of various casting procedures. The factors are rated A through E; letter A is the most advantageous.

Table 6. Comparison of Casting Methods^a

Factor	Casting method						
	Sand	Die	Invest-men t	Perma-nent mold	Plaster	Centri-fugal	Contín-uou s
tolerance	C	A	A	B	A	D	C
surface finish	D	B	A	C	A	D	C
thickness of section	C	D	D	C	D	A	A
pattern cost	A	E	B	C	B	A	A
ease of getting into production	A	D	B	C	B	C	C
production rate	B	A	C	C	D	B	C
cost per piece	B	A	C	C	D	A	C
flexibility as to alloy	A	D	A	D	C	A	A
limitation of size	A	D	C	B	D	A	A

^a Comparisons are all relative based on letters A through E; A is most advantageous.

Alloy Identification and Chemistry

The nominal chemical composition and identification of the most important copper casting alloys are listed in Table 7. These alloys are identified by name and by the Unified Numbering System. The use of names is not recommended.

Table 7. Nominal Composition of Some Casting Alloys

Common name	UNS designation	Nominal composition, wt %						
		Cu	Sn	Pb	Zn	Fe	Al	Others
copper	C 80100	99.95						0.05 ^a
chrome–copper	C 81500 ^b	98.0 ^c						1.0 Cr
beryllium–copper	C 81700 ^b	94.25 ^c						1.0 Ag 0.4 Be 0.9 Co 0.9 Ni
ASTM B22 high strength yellow brass	C 86300	63.0			25. 0	3.0	6.0	3.0 Mn
gun metal	C 90500	88.0	10. 0		2.0			
tin bronze 84:16	C 91100	84.0	16. 0					
tin bronze 81:19	C 91300	81.0	19. 0					
high leaded tin bronze	C 93700	80.0	10. 0	10. 0				
ASTM B61 steam bronze	C 92200	88.0	6.0	1.5	4.5			
ASTM B62 leaded red brass	C 83600	85.0	5.0	5.0	5.0			
ASTM B66 phosphorus bronze	C 94400	81.0	8.0	11. 0				0.35 P
high leaded tin bronze	C 93800	78.0	7.0	15. 0				
medium bronze	C 94500	73.0	7.0	20. 0				
high leaded tin bronze	C 94300	70.0	5.0	25. 0				
ASTM B67 journal bronze	C 94100	70.0	5.5	18. 0	3.0 ^a			
ASTM B148 aluminum bronze 9A	C 95200	88.0				3.0		9.0
aluminum bronze 9B	C 95300	89.0				1.0		10.0
aluminum bronze 9C	C 95400	85.0				4.0		11.0
aluminum bronze 9D	C 95500	81.0				4.0	11. 0	4.0 Ni
aluminum–silicon bronze	C 95600	91.0					7.0	2.0 Si
manganese–aluminum bronze	C 95700	75.0				3.0	8.0	12.0 Mn, 2.0 Ni
nickel–aluminum bronze	C 95800	81.0				4.0	9.0	1.0 Mn, 5.0 Ni
ASTM B176 die-casting yellow brass	C 85800	58.0	1.0	1.0	40. 0			
die-cast silicone brass	C 87800	82.0			14. 0			4.0 Si
silicon–yellow brass	C 87900	65.0			34. 0			1.0 Si
ASTM B584 leaded red brass	C 83800	83.0	4.0	6.0	7.0			
leaded semired brass	C 84400	81.0	3.0	7.0	9.0			
	C 84800	76.0	3.0	6.0	15. 0			
	C 85200	72.0	1.0	3.0	24. 0			
commercial no. 1 yellow brass	C 85400	67.0	1.0	3.0	29. 0			
yellow brass	C 85700	63.0	1.0	1.0	34. 7		0.3	
high strength yellow brass	C 86200 ^d	64.0			26. 0	3.0	4.0	3.0 Mn
leaded high strength yellow brass	C 86400	59.0		1.0	40. 0	2.0 ^a	1.5 ^a	1.5 ^a Mn
high strength yellow brass	C 86500	58.0	0.5		39. 5	1.0	1.0	1.5 ^a Mn
leaded high strength yellow brass	C 86700	58.0		1.0	41. 0	3.0 ^a	3.0 ^a	3.5 ^a Mn
silicon bronze	C 87200	89.0 ^c	1.0 ^a	0.5 ^a	5.0 ^a	2.5 ^a	1.5 ^a	1.5 ^a Mn, 4.0 Si
silicon brass	C 87400	83.0			14. 0			3.0 Si
silicon brass	C 87500	82.0			14. 0			4.0 Si
tin bronze	C 90300	88.0	8.0		4.0			
leaded tin bronze	C 92300	87.0	8.0	1.0 ^a	4.0			
high leaded tin bronze	C 93200	83.0	7.0	7.0	3.0			
	C 93500	85.0	5.0	9.0	1.0			

nickel–tin bronze	C 94700	88.0	5.0		2.0	5.0 Ni
leaded nickel–tin bronze	C 94800	87.0	5.0	1.0 ^a	2.5 ^a	5.0 Ni
	C 94900	80.0	5.0	5.0	5.0	5.0 Ni
12° nickel–silver	C 97300	56.0	2.0	10.	20.	12.0 Ni
				0	0	
15° nickel–silver	C 97400	59.0	3.0	5.0	16.	17.0 Ni
					0	
20° nickel–silver	C 97600	64.0	4.0	4.0	8.0	20.0 Ni
25° nickel–silver	C 97800	66.0	5.0	2.0	2.0	25.0 Ni
copper–nickel 90:10	C 96200	87.5				1.5
copper–nickel 70:30	C 96400	68.0				0.7
						10.0 Ni, 0.9 Mn
						30.0 Ni, 0.8 Mn

^a Maximum value.
^b Responds to heat treatment.
^c Minimum value.
^d Several compositions are available that meet mechanical property specifications.

Effect of Various Alloying Elements. The mechanical properties of cast copper alloys are a function of alloying elements and their concentrations. The specific effects of a number of these alloying elements are given in the following sections.

Zinc. Zinc [7740-66-6] is added to copper as a predominating alloying element in concentrations of 5–40° o, forming the alloy series known as the brasses. Zinc increases the tensile strength at a significant rate up to a concentration of ca 20° o, whereas the tensile strength increases only slightly more for additions of zinc of 20–40° o. The Rockwell F scale hardness is substantially an increasing straight-line function with zinc additions of 5–35° o. The higher zinc concentrations in high strength yellow brass alloys form a duplex alpha–beta structure.

Zinc up to 5° o is added to tin bronze alloys to tighten the structure and to act as a deoxidizer, thereby aiding in producing sound castings for pressure applications. In yellow brasses zinc can impart a freedom from gas porosity because, as the melt is heated until the zinc boils, the zinc vapor sweeps the melt free of gas.

Zinc is not considered a very harmful impurity to most alloys. Cast copper–zinc alloys are described as red brasses and leaded red brasses, semired, silicon, yellow, and high strength yellow brasses. Red brasses contain zinc as the principal alloying element along with some tin and lead. Semired brasses contain less copper than do red brasses. Yellow brasses contain zinc as the principal alloying element accompanied by small amounts of tin and lead or other designated elements. High strength yellow brasses are alloys that contain zinc with smaller amounts of iron, aluminum, nickel, and lead.

Tin. Tin [2440-31-5] added to copper in concentrations of 5–20° o forms the tin–bronze alloy series. Leaded tin bronze is also produced. Typically a deoxidizer is added to the melt to produce a clean structure. Phosphorus is the usual element employed as a deoxidizer. Zinc may also be used, but it is not effective as a deoxidizer until it is present in concentrations of ca 2° o. Delta tin eutectoid develops in some castings containing as little as 6–8° o tin because of nonequilibrium freezing conditions. The cooper–tin constitution diagram under equilibrium condition indicates delta tin should not form until the tin concentration exceeds 16° o at 520°C. Tin imparts strength and hardness to copper-base alloys, making them tough and wear resistant. Worm gears are often made from tin–bronze containing 8° o or more tin. Tin also enhances the corrosion resistance of copper-base alloys in nonoxidizing media.

Small amounts of tin (3–5° o) are added to leaded red brass and semired brasses to increase the strength and hardness of alloys. For example, alloy UNS C 94700 (88° o Cu, 5° o Sn, 5° o Ni, and 2° o Zn) deoxidized with phosphorus, is heat-treatable to provide high strength.

Tin is usually not regarded as an impurity except in high tensile strength manganese–bronze where it is limited to 0.2° o maximum. Tin lowers both the tensile strength and the ductility of the alloy.

Lead. Lead [7439-92-1] is added to copper in amounts up to 40° o. Lead is insoluble in copper-base alloys, and because of its low melting point it is found distributed in the grain boundaries of the casting. Because it imparts a certain degree of brittleness to the structure, it enhances machining operations by causing the alloy to break into chips as cutting tools are thrust into the matrix. Nonleaded alloys usually machine with the formation of long curls rather than chips, and the surface is not as smooth. Additions of lead up to 1.5° o significantly improve machinability without a serious decrease in tensile strength. Lead concentrations of 5–25° o greatly increase machinability. Alloys with lead concentrations equal to or greater than the tin content are used for bearing applications requiring resistance to both wear and friction. Lead added to copper in amounts of about 35–40° o forms a useful bearing alloy.

Lead is considered undesirable in high strength manganese bronze, silicon bronze, and silicon brass. It affects the surface of silicon bronze and silicon brass, causing noticeable darkening and pockmarking.

Silicon. Silicon [7740-21-3] added to copper forms alloys of high strength and toughness along with improved corrosion resistance, particularly in acidic media. Silicon in small amounts can improve fluidity. Silicon is a very harmful impurity in leaded tin bronze alloys, however, because it contributes to lead sweat and unsoundness.

Aluminum. Aluminum [7429-90-5] added as the predominating alloying element to copper forms a series of high strength alloys called aluminum bronzes. Aluminum forms solid solutions with copper up to about 9.5° o. High strength yellow brasses contain aluminum in varying amounts. Some of the aluminum-bearing alloys form a second phase (beta) and are heat-treatable. With still larger (>12%) amounts of aluminum, another extremely hard phase, gamma-2, forms.

Aluminum present in leaded tin bronze alloys promotes unsoundness.

Iron. Iron [7439-89-6] added to copper alloys adds strength to the silicon, aluminum, and manganese bronzes. It combines with aluminum or manganese to form hard, intermetallic compounds. The hard compounds embedded in the matrix, which have high melting points and act as a grain refiner, increase the alloy's wear-resistance qualities. Undissolved iron in the alloy leads to nonuniform hardness and interferes with machining.

Phosphorus. Phosphorus [7723-14-0] is used principally as a deoxidizer in copper and high copper alloys. The alloy should contain a minimum residual of 0.02° o phosphorus to ensure complete deoxidization. Lesser amounts of residual phosphorus can form an equilibrium system with copper and oxygen. Such castings are subject to damage (embrittlement) when heated in a hydrogen atmosphere, depending on temperature and hydrogen concentration of the environment.

Boron. Boron [7440-42-8] is a commercial deoxidizer.

Manganese. Manganese [7439-96-5] is added as an alloying element in high strength brasses, where it forms compounds with other elements such as iron and aluminum. Manganese may also be used as a deoxidizer, although that is not a common usage.

Nickel. Nickel [7740-02-0] added to copper markedly whitens the resulting alloy. The element added to copper at the rate of 10–30° o produces the so-called cupronickel alloys, which have very high corrosion resistance. Iron, up to a nominal 1.4° o, added along with nickel significantly enhances the resistance toward cavitation or impingement corrosion. Added to bronzes, nickel refines the cast grain structure and adds toughness. Nickel improves strength and corrosion resistance. An alloy series containing 10–25° o nickel along with tin, lead, and zinc as principal alloy elements is known as the nickel–silvers. Nickel, also added to some of the high tin–gear bronze alloys to enhance wear properties, does not have deleterious effects as an impurity, and many specifications allow about 1° o.

Beryllium. Beryllium [7440-41-7] added to copper forms a series of age- or precipitation-hardenable alloys. These heat-treatable alloys are the strongest of all known copper-base alloys (see COPPER ALLOYS, WROUGHT COPPER ALLOYS).

Chromium. Chromium [7440-47-3] also forms heat-treatable copper alloys. These alloys, in the heat-treated condition, have a Brinell hardness of about 120 and an electrical conductivity of about 80° o IACS.

Arsenic, Antimony, and Phosphorus. Arsenic [7440-38-2], antimony [7440-36-0], and phosphorus can be added in small quantities

(0.05% α) to the all α -phase brass alloys containing less than 80% α copper to inhibit the dezincification type of corrosion in yellow brass alloys.

Economic Aspects

Casting is used for irregular external and internal shapes that are impractical, impossible, or too costly to produce using other methods. The choice of an alloy for any casting usually depends on four factors: metal cost, castability, properties, and final cost. A cost analysis determines the most economical method of producing a casting, although frequently the choice can be based on experience.

Metal cost, a minor consideration if only a few castings are made, becomes of prime importance when large amounts of metal are required, the casting is needed in large quantities or is a competitive item, or when metal cost is a significant factor in the final cost. Final cost must be considered, because the initial advantage of a lower metal cost may be eliminated by an increasing cost resulting from an inherent property of the alloy. One such property is castability.

Castability is not to be confused with fluidity, which is the ability of a molten alloy to fill a mold cavity in every detail. Castability is the ease with which an alloy responds to ordinary foundry practice without undue attention to gating, risering, melting, sand conditions, and any other factors involved in producing sound castings.

Health and Safety Factors

During melting and pouring, certain metals, such as zinc, volatilize, enter the atmosphere, and immediately oxidize to solid particulates, forming a smoke. Some of these fumes or smokes can be hazardous to health, depending on the chemical composition of the particulate, its concentration in the off-gas, and the duration of exposure. More and more melt and casting shops are significantly controlling the atmospheric conditions in working areas by the use of adequate ventilating hoods, ducts, and exhaust fans in strategic locations. Rather than being exhausted into the atmosphere outside the operating plant, the particulates can be captured in bag-filtering and automatic collection systems (see AIR POLLUTION CONTROL METHODS). Care must be exercised in the design and operation of such devices to prevent hot gases or sparks from igniting the bag-filter material. Fires in the collecting system can be very costly. Gas temperatures are controlled by the length of the ducts between the heat source and the filter and also by the volume of gas being moved through the system. Often a cyclone duct collector is placed in the system just before the baghouse for the purpose of collecting sparks from smoldering or burning carbonaceous material to prevent them from contacting the filters in the baghouse. Workers may also wear masks.

Local and national agencies have set limits on the quantities of particulates permissible in casting-shop atmospheres and in the exhaust from collection systems. Agencies in the United States are the EPA, NIOSH, and OSHA.

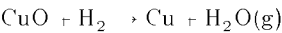
Particulates collected in foundry atmospheres have been found to contain compounds of zinc, copper, lead, iron, aluminum, magnesium, and silicon as well as carbon and oily material. The specific composition and concentration of any smoke depends on the composition of the material being melted and on the melting and pouring practice.

Foundry Practice

Copper and High Copper Alloys. Copper castings contain a minimum of 99.7% copper. High copper alloys contain more than 94% copper. The coppers are generally specified to have high electrical conductivity and therefore must be of high purity. It is usually necessary to start with ingots of fire-refined or electrolytic-grade copper, although high grade or number one copper scrap of known chemistry may also be used (see RECYCLING, NONFERROUS METALS.)

Copper is susceptible to gassing and must therefore be protected from direct flame contact during melting. Copper melts can be contaminated with either hydrogen or oxygen. Hydrogen is soluble in the melt and oxygen forms copper oxide (see COPPER COMPOUNDS). Melts should be protected from air by the use of a melt cover such as dry charcoal or carbon pieces. A cover decreases the area of the melt exposed to air and also reduces the amount of fuming.

Copper(II) oxide [1317-38-0] can also cause porosity in the finished casting by combining with hydrogen formed by the dissociation of water in the mold material to form steam within the melt, thus causing holes during solidification.



If no cover is used, copper should be melted under a slightly oxidizing atmosphere to reduce possible hydrogen solution. The oxides floating on the melt then tend to become the cover.

Copper melt heated to a temperature 30–55°C above its melting point should be deoxidized using calcium boride [1200-99-7], CaB , lithium [7439-93-2], or copper–phosphorus 15% alloy. The deoxidizer is plunged beneath the melt surface. A shrink test may be made by pouring a test piece ca 4 cm in diameter by 10–13 cm in length into a mold of the same type to be used for the cast pieces. If the test piece shrinks, the melt is considered ready to pour. Experience is generally a good guide.

If the test piece contains many holes, gas is indicated and the melt may be flushed with nitrogen to control this situation. Small amounts of deoxidizer should be added to the gas-free melt until shrinking of the test piece occurs.

Cores should be made with a minimum amount of binder or volatile compounds and should also be readily collapsible because copper is susceptible to hot tearing.

Deoxidized copper is a high shrinkage material, and castings must be suitably risered to provide proper feeding. Exothermic compounds, such as iron oxide and magnesium, may be added to the top of the riser to keep the copper hot and molten to aid in feeding. Insulation and exothermic riser sleeves may also be used.

Contamination of the melt with elements that remain in solid solution must be avoided, and excess deoxidizer must be controlled to obtain castings of high electrical conductivity. Properties are shown in Table 8.

Table 8. Properties of Copper and High Copper Alloys

Property	Copper C 80100	Chromium–copper C 81500	Beryllium–copper C 81700
melting point, °C	1064–1083	105–1085	1029–1068
pouring temperature, °C			
light castings	1149–1204	1149–1204	1132–1188
heavy castings	1110–1166	1110–1165	1093–1149
specific gravity	8.94	8.82	8.75
dross generation	very low	high	low
gassing	high	moderate	moderate
cast yield	low	low	moderate
pattern maker's shrinkage, cm /m	1.56	2.08	1.56

Coppers and high copper alloys can be successfully cast using the centrifugal, continuous, investment, permanent, plaster, and sand molding methods.

Uses. Copper and high copper alloys are typically used as electrical and thermal conductors. UNS C 80100 is corrosion and oxidation resistant;

UNS C 81500 (chromium–copper alloy) is used structurally where strength and hardness are required; and UNS C 81700 (beryllium–copper alloy) is used structurally where high strength and hardness are required.

Red Brass Alloys. In forming red brass alloys, which include leaded red and leaded semired brasses, caution should be exercised to prevent gas absorption by flame impingement or the melting of oily scrap, or metal loss through excessive oxidation of the melt surface. To prevent excessive zinc volatilization, the melt must be poured as soon as it reaches the proper temperature. The melt should be finally deoxidized and cast at ca 1065–1230°C as measured with a pyrometer. Fluxing is usually not needed if clean material has been melted.

The deoxidizer (copper–phosphorus 15^o alloy) is added just before pouring in a quantity calculated to result in less than 0.01^o residual phosphorus. The suggested addition is ca 1 g kg of melt. Zinc-containing scrap should be melted first, and then any ingot copper along with zinc necessary to replace any burned out during melting. The amount of make-up zinc to be added varies with practice and is judged by experience. Care should be taken not to heat the melt more than 83°C above the pouring temperature. Proper gating and rising is very important to compensate for the alloy shrinkage that occurs. Directional solidification techniques must be employed to assure the integrity of the casting. Chills are often utilized in the mold to assist the solidification direction.

Properties of red brass alloys are given in Table 9. The members of this group are cast using the centrifugal, continuous, investment, and sand molding methods. General tensile strengths vary from 170 to 210 MPa (25,000–30,000 psi) minimum as cast in sand molds.

Table 9. Properties of Red Brass Alloys

Property	Leaded red brass		Leaded semired brass	
	C 83600	C 83800	C 84400	C 84800
melting point, °C	854–1010	843–1004	843–1060	832–954
pouring temperature, °C				
light castings	1149–1288	1149–1266	1149–1260	1149–1260
heavy castings	1066–1177	1066–1177	1066–1177	1066–1177
specific gravity	8.7–8.88	8.6–8.7	8.6–8.8	8.55–8.70
dross generation	low	low	moderate	moderate
gassing	moderate	moderate	moderate	moderate
cast yield	high	high	high	high
pattern maker's shrinkage, cm m	1.56	1.56	1.56	1.56

Uses. Leaded red brass alloys are used in plumbing goods, in fittings and small gears (UNS C 83600), and in low pressure valves and fittings and general hardware (UNS C 83800). Leaded semired brasses are used in general hardware, low pressure valves, and fittings; ornamental casting and plumbing goods (UNS C 84400); and in cocks and faucets (UNS C 84800).

Tin Bronze Alloys. Tin bronze alloys are successfully melted in any type of foundry furnace. Best results are obtained by protecting the melt from either direct flame contact or excessive oxidation. Melting should be carried out in a slightly oxidizing atmosphere. Rapid melting is desired to reduce the opportunity of gas absorption in the melt. Fracture tests exhibit an open grain and discoloration when the melt has been exposed to reducing conditions, contaminants such as Al and Si, or excessive superheating. Oily scrap should be avoided. Pouring temperatures range from 1010 to 1260°C and should be measured with a pyrometer. Pouring should take place at a temperature that gives the best results for a particular gating or rising system. The temperature should be high enough to avoid internal shrinkage that can lead to voids, because gating and rising may not always completely eliminate such internal unsoundness. Directional solidification needs to be planned as part of the gating system.

Alloy scrap should be melted first and then have any copper added to the melt. It is general practice to deoxidize the melt before pouring. The alloys may be degassed by superheating ca 40°C to boil the zinc and allowing it to quietly air cool before pouring. Zinc must be added to replace the amount boiled off. Flushing the melt with dry nitrogen also degasses the melt. Lost zinc must be replaced. Gating and rising of castings of alloys is very important, and careful consideration should be exercised for each configuration. Sound castings are made by following fundamental foundry techniques.

Properties of copper–tin–lead alloys are listed in Table 10. The members of the tin bronze alloy group are cast using the centrifugal, continuous, permanent, plaster, and sand molding methods. Leaded tin–bronze alloys have minimum tensile strengths of 234–248 MPa (34,000–36,000 psi) as cast in sand molds, whereas the minimum tensile strengths for high leaded tin–bronze alloys are 138–207 MPa (20,000–30,000). The values are based on measurement of test bars cast in sand molds.

Table 10. Properties of Tin-Bronze Alloys

Property	Gun metal C 90500	Tin bronze 84:16 C 91100	Tin bronze 81:19 C 91300	Steam bronze C 92200	Leaded tin bronze C 92300	High leaded tin bronze				
						C 93200	C 93500	C 93700	C 93800	C 94300
melting point, °C	854–999	818–960	818–889	826–988	854–999	954–982	854–999	762–929	854–954	899–927
pour t, °C										
light castings	1093–126 0	1093–123 2	1093–120 4	1093–126 0	1149–126 0	1093–123 2	1038–120 4	1093–123 2	1093–123 2	1093–120 4
heavy castings	1038–114 9	1038–112 1	1038–109 3	1038–114 9	1049–114 9	1038–112 1	1038–120 4	1010–114 9	1038–114 9	1010–109 3
specific gravity	8.7			8.6–8.8	8.7–8.87	8.85–8.9	8.87	8.9–9.1	9.1–9.4	9.2–9.5
dross generation	low	low	low	low	low	low	low	low	low	low
gassing	moderate	moderate to high	moderate to high	moderate	moderate	moderate	moderate	moderate	moderate	moderate to high
cast yield shrinkage, ^a cm m	moderate 1.56	moderate 1.56	moderate 1.56	moderate 1.56–1.82	moderate 1.56–1.82	high 1.56–1.82	high 1.56	high 1.04–1.56	high 1.56	high 1.56

^a Pattern maker's shrinkage.

Uses. Leaded tin bronzes are used as valves and as fittings for pressure parts (UNS C 92200) and for high pressure steam (UNS C 92300). The high leaded tin bronze alloys are used as bearings and bushings (UNS C 93200); as small bearings, bushings, and backing for babbitt-lined bearings (UNS

C 93500); as high speed and high pressure bearings (UNS C 93700); as general service, moderate pressure, acid mine water exposure bearings (UNS C 93800); and as high speed and light load bearings (UNS C 94300).

Manganese Bronze Alloys. The melt should be heated to the highest temperature without causing excessive zinc boil or flare, and pouring temperature is important enough to be measured. Zinc losses can substantially change the composition of the casting. Manganese bronze alloys are dross formers and therefore tend to develop their own melt covers. However, if open flame melting is practiced, a flux cover can be used. Because the alloys form dross (oxidation of zinc, aluminum, and manganese) readily, choke gates, strainer cores, and pouring basins should be arranged to cause the melt to enter the mold with as little turbulence as possible.

Manganese bronze alloys have high shrink characteristics, and special attention must be given to gating and rising systems for individual castings. Hot tops and insulated sleeves are often used in casting manganese bronze because of an insulating effect on the riser, thus promoting better feeding to the casting. Good castings can be produced if the composition is controlled and good melting, gating, and rising practices are followed.

The alloys in this series have been cast using centrifugal, continuous, die, investment, permanent, plaster, and sand molding systems. Table 11 lists properties and characteristics of manganese–bronze alloys. The minimum tensile strength for leaded high strength yellow brass is 413 MPa (60,000 psi) for sand-cast test bars, and the minimum tensile strength for high strength yellow brasses varies from 448 to 758 MPa (65,000–110,000 psi).

Table 11. Properties of High Strength Yellow Brass^a

Property	Leaded C 86400	Nonleaded		
		C 86200	C 86500	C 86300
melting point, °C	860–880	899–927	862–880	885–923
pouring temperature, °C				
light castings	1038–1121	1066–1177	1038–1093	1066–1177
heavy castings	954–1038	982–1066	954–1038	982–1066
specific gravity	8.0–8.4	7.9	8.0–8.5	7.7–8.0
dross generation	high	high	high	high
gassing	low	low	low	low
cast yield	low	low	low	low
pattern maker's shrinkage, cm m	1.82–2.08	2.08–2.60	1.82–2.34	2.08–2.60

^a Manganese–bronze alloys.

Uses. The general uses for the manganese alloy group are alloy UNS C 86200, as marine fittings, gears, and gun mounts; alloy UNS C 86300 as bearings and bushings; and alloy UNS C 86400, leaded manganese bronze, as free-machining manganese bronze, valve stems, marine fittings, light duty gears, and propellers.

Aluminum Bronze Alloys. The melting precautions for aluminum bronzes are basically the same as those for the other copper-base alloys. The danger of gas absorption is not as great as with tin bronze alloys because of the protective film of aluminum oxide on the surface of the melt. Nevertheless, the alloys do dissolve gas and therefore precautions must be taken to prevent this. Although pouring temperatures are not as critical, temperature control is needed. The pouring range is 1065–1260°C, depending on casting size and section; ladle surfaces should be kept well skimmed during pouring, and the melt should be exposed to air as little as possible. The shortest distance from the pouring ladle to the mold should be used to hold turbulence to an absolute minimum. Every precaution to eliminate dross should be taken.

The gating and rising system for cast aluminum bronze is extremely important and must be arranged to introduce the metal quietly at the lowest portion of the mold. The alloys shrink well; hence the gating and rising must be well adapted to the particular casting. See Table 12 for properties of these alloys. Alloys C 95300, C 95400, and C 95500 are heat-treatable for increased mechanical properties and the last two should be temper-annealed if used in a corrosive environment.

Table 12. Properties of Aluminum Bronze Alloys

Property	C 95200	C 95300	C 95400	C 95500
melting point, °C	1038–1043	1038–1052	1027–1038	1038–1054
pouring temperature, °C				
light castings	1121–1204	1121–1204	1260	1232
heavy castings	1093–1149	1066–1149	1777	1066
specific gravity	7.3–7.5	7.3–7.65	7.50	7.49–7.66
dross generation	high	high	high	high
gassing	moderate	moderate	moderate	moderate
shrinkage, ^a cm m	2.60–2.86	1.82–2.60	1.82–2.60	2.08

^a Pattern maker's shrinkage.

Aluminum bronze alloys have been successfully cast in the centrifugal, continuous, permanent, plaster, and sand molding methods. Depending on the alloy, the minimum tensile strengths of sand-cast test bars are 448–620 MPa (65,000–90,000 psi).

Uses. Aluminum bronze alloy UNS C 95200 is used as acid-resistant pumps, pump rods, bushings, and gears; alloy UNS C 95300, as pickling baskets, gears, and marine equipment; alloy UNS C 95400, as bearings, gears, valve seats, valve guides, and pickling hooks; and alloy UNS C 95500, as valve guides and seats, corrosion-resistant parts, bushings, gears, and worms.

Silicon Bronze and Silicon Brass Alloys. For silicon-containing alloys best casting results are obtained if the melting is accomplished using slightly oxidizing conditions. Rapid heating to the melting point is required to control dissolved gases in the melt. Superheating the melt to ca 85°C above the pouring temperature is desired. Undisturbed cooling to the pouring temperature allows the dross to float on the melt surface. Both melting and pouring temperatures are important and should be measured. Pouring temperatures for these alloys are 1040–1175°C, depending on size and section thickness of the casting. Prolonged holding of the melt at pouring temperature should be avoided. Molds should be ready to receive the melt as soon as it is at the proper temperature. Melting oily scrap should be avoided. Silicon is a powerful deoxidizing element and can easily be lost from the melt in the form of dross.

Covers for the melt can be used. Preburned charcoal has been successfully used as a cover and is applied hot directly from the preburner. Fluxing and deoxidation is not necessary because of the strong affinity silicon has for oxygen. Silicon-bearing scrap must be kept segregated from other scrap, because silicon as an impurity in other alloys can promote a very coarse dendritic structure and weak porous castings.

If gas has been absorbed, some degassing treatment is necessary. The melt may be flushed using dry nitrogen or a degassing flux may be used. Copper–silicon alloys shrink less than manganese bronze and aluminum bronze but more than tin bronze. Good gating and rising systems are required. The properties of these alloys are listed in Table 13.

Table 13. Properties of Silicon Bronze and Silicon Brass Alloys

Property	Silicon-bronze C 87200	Silicon-brass	
		C 87400	C 87500
melting point, °C	860–971	821–916	821–917
pouring temperature, °C			
light castings	1093–1177	1093–1177	1093–1177
heavy castings	1038–1066	1038–1066	1038–1066
specific gravity	8.30–8.44	8.30–8.44	8.30–8.44
dross generation	low	low	low
gassing	high	moderate to high	moderate to high
casting yield	moderate	moderate	moderate
pattern maker's shrinkage, cm m	1.82	1.82	1.82

The alloys in the copper–silicon group have been cast using centrifugal, investment, die, permanent, plaster, and sand molding methods. The minimum tensile strengths for sand-cast test bars are 310–413 MPa (45,000–60,000 psi).

Uses. Uses for copper–silicon alloys are silicon bronze, UNS C 87200, as bearings, pumps, valve parts, marine fittings, and corrosion-resistant castings; silicon brass UNS C 87400 as bearings, gears, impellers, valve stems, and clamps; and silicon brass UNS C 87500 as small propellers, valve stems, gears, and bearings.

Copper–Nickel and Leaded Nickel Bronze and Brass Alloys. High pouring temperatures cause most casting troubles with this group. The temperatures make a sand mold of low permeability practically unusable because the steam generated may exceed the venting power of the sand. Gases dissolve more readily at the higher temperatures and should be removed before pouring the melt. Degasifiers are used as a matter of standard practice. Copper–nickel alloys should be melted under slightly oxidizing conditions. Scrap must be clean. Cutting oils or compounds or other organic matter must be removed before melting to control dissolved gases in the melt that can lead to inferior castings.

Time must be allowed for the degasifier to be effective, and the treated melt should be tested with a so-called pitch test to determine the condition of the melt.

An accurate and rapid method for evaluating the gas content of melts has been developed. The method consists of solidifying a sample of the melt in a vacuum and measuring the degree of swelling in the solidified sample. Melting and pouring temperatures must be controlled by measurement and the melt poured at 1095–1425°C, depending on the alloy, casting size, and section thickness. Gating and rising are especially important, and careful consideration of the particular casting must be made.

Properties of copper–nickel alloys are listed in Table 14. The alloys in the copper–nickel group have been successfully cast using the centrifugal, investment, permanent, and sand molding methods. The minimum tensile strengths on test bars cast in sand molds are 207–310 MPa (30,000–45,000 psi).

Table 14. Properties of Copper–Nickel Alloys and Leaded Nickel Bronze and Brass

Property	Copper–nickel		Leaded nickel bronze		Leaded nickel brass C 97300
	C 96200	C 96400	C 97600	C 97800	
melting point, °C	1099–1149	1171–1238	1108–1143	1140–1180	1010–1040
pouring temperature, °C					
light castings	1316–1427	1371–1483	1260–1427	1316–1427	1204–1316
heavy castings	1204–1316	1316–1399	1232–1316	1260–1316	1993–1204
specific gravity	8.94	8.94	8.8–8.9	8.8–8.9	8.9–8.95
dross generation	low	moderate	moderate to high	moderate to high	high
gassing	high	high	moderate to high	moderate to high	moderate
cast yield	low	low	moderate	moderate	moderate
pattern maker's shrinkage, cm m	1.56	1.82	1.56	1.56	1.04–1.56

Uses. Copper–nickel–iron alloys, UNS C 96200 (90:10 copper:nickel) and UNS C 96400 (70:30 copper:nickel), are used in corrosion-resistant marine (seawater) applications. UNS C 96400 is used for corrosion-resistant marine elbows, flanges, valves, and pumps. Leaded nickel–brass, UNS C 97300 (12% nickel–silver), is used for hardware fittings, valves, and statuary and ornamental castings.

The leaded nickel bronzes, UNS 97600 (20% nickel–silver) and UNS C 97800 (25% nickel–silver), are used for marine, ornamental, and sanitary castings, and valves and pumps, and for ornamental hardware, sanitary castings, valves, and musical instrument parts, respectively.

BIBLIOGRAPHY

"Cast Copper Alloys" under "Copper Alloys" in *ECT* 1st ed., Vol. 1, pp. 458–467, by G. P. Halliwell, H. Kramer & Co.; in *ECT* 2nd ed., Vol. 6, pp. 244–265, by R. E. Ricksecker, Chase Brass & Copper Co.; in *ECT* 3rd ed., Vol. 7, pp. 69–96, R. E. Ricksecker, Chase Brass & Copper, Inc.

1. *Casting Copper-Base Alloys*, American Foundrymen's Society, Des Plaines, Ill., 1984, pp. 232–233.

General References

P. R. Beeley, *Foundry Technology*, Halsted Press, New York, 1972.

Casting Defects Handbook, American Foundrymen's Society, Des Plaines, Ill., 1972.

Casting Design, American Foundrymen's Society, Des Plaines, Ill., 1987.

G. J. Cook, *Engineered Castings*, McGraw-Hill Book Co., New York, 1961.

Copper and Copper Alloys, ASTM Standards Book, vol. 2.01, American Society for Testing Materials, Philadelphia, 1989.

F. Dunlevey, "Casting Beryllium–Copper Alloys," *Modern Casting Magazine* (Nov. 1991).

C. J. Evans, *Foundry Trade J.*, 143 (Dec. 22, 1977).

Gating and Rising of Copper-Base Alloys, American Foundrymen's Society, Des Plaines, Ill., 1986.

R. W. Heine and P. C. Rosenthal, *Principles of Metal Casting*, McGraw-Hill Book Co., Inc., New York, 1955.

A. W. Hudd, "Development of Continuous Casting of Copper Slab and Billet at IMI Refiners Limited," *Metall. Mater. Technol.* **9**(11), (1977).

Metals handbook, 9th ed., vol. 2, American Society for Metals, Metals Park, Ohio, 1979.

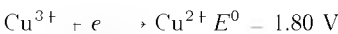
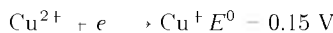
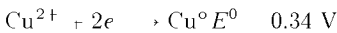
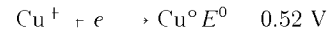
J. L. Morris, *Metal Castings*, Prentice-Hall Inc., Englewood Cliffs, N.J., 1957.

D. G. Schmidt, "The Challenge of Copper Alloy Castings in the 21st Century," *Transactions AFS*, 1991.
I. D. Simpson and N. Standish, *Metallography* **10**(4), 433 (Oct. 1977).
Unified Numbering System for Metals and Alloys, SAE HS 1086a or ASTM DS-56a, 2nd ed., 1977.

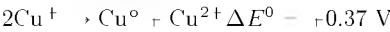
Daniel L. Twarog
American Foundrymen's Society

COPPER COMPOUNDS

Copper compounds, which represent only a small percentage of all copper production, play key roles in both industry and the biosphere. Copper [7440-50-8], mol wt = 63.546, [Ar]3d¹⁰4s¹, is a member of the first transition series and much of its chemistry is associated with the copper(II) ion [15158-11-9], [Ar]3d⁹. Copper forms compounds of commercial interest in the +1 and +2 oxidation states. The standard reduction potentials, E⁰, for the reasonably attainable valence states of copper are



The copper(I) ion [17493-86-6] disproportionates spontaneously, pK = -5.95, in aqueous solution according to



The concentration of copper(I) ion remaining in solution is not appreciable. However, aqueous copper(I) ion can be stabilized by complex formation with various agents such as chloride, ammonia, cyanide, or acetonitrile.

The tri- or tetraamine complex of copper(I), prepared by reduction of the copper(II) tetraamine complex with copper metal, is quite stable in the absence of air. If the solution is acidified with a noncomplexing acid, the formation of copper metal, and copper(II) ion, is immediate. If hydrochloric acid is used for the neutralization of the ammonia, the insoluble cuprous chloride [7758-89-6], CuCl, is precipitated initially, followed by formation of the soluble ions [CuCl₃]²⁻, [CuCl₄]³⁻, and [CuCl₅]⁴⁻ as acid is increased in the system.

The copper(I) ion, electronic structure [Ar]3d¹⁰, is diamagnetic and colorless. Certain compounds such as cuprous oxide [1317-39-1] or cuprous sulfide [22205-45-4] are intensely colored, however, because of metal-to-ligand charge-transfer bands. Copper(I) is isoelectronic with zinc(II) and has similar stereochemistry. The preferred configuration is tetrahedral. Linear and trigonal planar structures are not uncommon, in part because the stereochemistry about the metal is determined by steric as well as electronic requirements of the ligands (see COORDINATION COMPOUNDS).

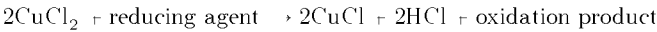
The stereochemical preference of the copper(II) ion is square planar or distorted octahedral because of the ligand field stabilization that arises from the 3d⁹ electronic configuration. This perturbation in an octahedral symmetry is known as Jahn-Teller distortion. Other configurations that occur for copper(II) include distorted tetrahedrons as well as a variety of five coordinate structures. Most copper(II) compounds are blue or green in color and exhibit a variety of magnetic phenomena (1). The majority of copper(II) compounds exhibit paramagnetic behavior as a result of the single unpaired 3d electron. There are, however, a significant number of polynuclear copper compounds that are sufficiently condensed to show spin-spin coupling of the unpaired electrons. This spin pairing may be so weak as to be observed only at near absolute zero temperatures or it may be strong enough to render the compound diamagnetic at room temperature or above. There have been reports of ferromagnetic polynuclear compounds as well. Probably the most significant, has been the high temperature superconductivity of copper oxide-containing materials (2,3).

Manufacture of Commercially Important Compounds

Copper(II) Carbonate Hydroxide. Basic copper carbonate, also named copper(II) carbonate hydroxide [12069-69-1], occurs in nature as the green monoclinic mineral malachite. The approximate stoichiometry is CuCO₃·Cu(OH)₂. There are two grades available commercially, the light and the dense. The light grade is produced by adding a copper salt solution to a concentrated solution of sodium carbonate, usually at 45–65°C. The blue, voluminous azurite [12070-39-2], C₂H₂Cu₃O₈, forms initially and converts to the green malachite within two hours. The dense product can be produced by boiling an ammoniacal solution to copper(II) carbonate (4) or by addition of a copper salt solution to sodium bicarbonate at 45–65°C. A dense product can also be produced by simultaneous addition of copper(II) salt solutions and soda ash solutions at controlled pH. Pure CuCO₃ has not been isolated.

Basic copper carbonate is essentially insoluble in water, but dissolves in aqueous ammonia or alkali metal cyanide solutions. It dissolves readily in mineral acids and warm acetic acid to form the corresponding salt solution.

Copper Chloride. Copper(I) chloride, CuCl, is a colorless or gray cubic crystal and occurs in nature as the mineral nantokite [14708-85-1]. The commercial product is white to gray to brown to green and of variable purity. Copper(I) chloride is usually produced at 450–900°C by direct combination of copper metal and chloride gas to yield a molten product (5–8). Once the reaction is initiated by heat it is self-sustaining and must be cooled. The molten product is variously cast, pilled, flaked, or ground depending on final use. Copper(I) chloride can be produced hydrometallurgically by reduction of copper(II) in the presence of chloride (9):

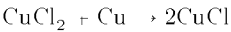


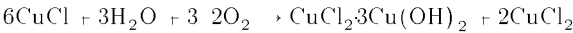
where the reducing agent can be sulfite, metallic copper, phosphorus acid, hydroxylamine, or zinc (10).

Copper(I) chloride is insoluble to slightly soluble in water. Solubility values between 0.001 and 0.1 g/L have been reported. Hot water hydrolyzes the material to copper(I) oxide. CuCl is insoluble in dilute sulfuric and nitric acids, but forms solutions of complex compounds with hydrochloric acid, ammonia, and alkali halide. Copper(I) chloride is fairly stable in air at relative humidities of less than 50%, but quickly decomposes in the presence of air and moisture.

Cupric chloride or copper(II) chloride [7447-39-4], CuCl₂, is usually prepared by dehydration of the dihydrate at 120°C. The anhydrous product is a deliquescent, monoclinic yellow crystal that forms the blue-green orthorhombic, bipyramidal dihydrate in moist air. Both products are available commercially. The dihydrate can be prepared by reaction of copper carbonate, hydroxide, or oxide and hydrochloric acid followed by crystallization. The commercial preparation uses a tower packed with copper. An aqueous solution of copper(II) chloride is circulated through the tower and chlorine gas is sparged into the bottom of the tower to effect oxidation of the copper metal. Hydrochloric acid or hydrogen chloride is used to prevent hydrolysis of the copper(II) (11,12). Copper(II) chloride is very soluble in water and soluble in methanol, ethanol, and acetone.

Copper(II) oxychloride [1332-65-6], Cu₂Cl(OH)₂, is found in nature as the green hexagonal paratacamite [12186-00-4] or rhombic atacamite [1306-85-0]. It is usually precipitated by air oxidation of a concentrated sodium chloride solution of copper(I) chloride (13–15). Often the solution is circulated through a packed tower of copper metal, heated to 60–90°C, and aerated.



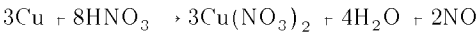


The mother liquor is separated from the product and returned to the tower. Copper(II) oxychloride is insoluble in water, but dissolves readily in mineral acids or warm acetic acid. The product dissolves in ammonia and alkali cyanide solution upon the formation of coordination complexes.

Copper Hydroxide. Copper(II) hydroxide [20427-59-2], Cu(OH)₂, produced by reaction of a copper salt solution and sodium hydroxide, is a blue, gelatinous, voluminous precipitate of limited stability. The thermodynamically unstable copper hydroxide can be kinetically stabilized by a suitable production method. Usually ammonia or phosphates are incorporated into the hydroxide to produce a color-stable product. The ammonia processed copper hydroxide (16–19) is almost stoichiometric and copper content as high as 64% is not uncommon. The phosphate produced material (20,21) is lower in copper (57–59%) and has a finer particle size and higher surface area than the ammonia processed hydroxide. Other methods of production generally rely on the formation of an insoluble copper precursor prior to the formation of the hydroxide (22–26).

Copper hydroxide is almost insoluble in water (3 μg/L) but readily dissolves in mineral acids and ammonia forming salt solutions or copper ammine complexes. The hydroxide is somewhat amphoteric dissolving in excess sodium hydroxide solution to form trihydroxycuprate [37830-77-6], [Cu(OH)₃][–], and tetrahydroxycuprate [17949-75-6], [Cu(OH)₄][–].

Copper Nitrates. The trihydrate [10031-43-3] crystallizes as blue rhombic plates. Copper(II) nitrate hexahydrate [13478-38-1], Cu(NO₃)₂·6H₂O, is produced by crystallization from solutions below the transition point of 26.4°C. A basic copper nitrate [12158-75-7], Cu₂(NO₃)(OH)₃, rather than the anhydrous product is produced on dehydration of the hydrated salts. The most common commercial forms for copper nitrate are the trihydrate and solutions containing about 14% copper. Copper nitrate can be prepared by dissolution of the carbonate, hydroxide, or oxides in nitric acid. Nitric acid vigorously attacks copper metal to give the nitrate and evolution of nitrogen oxides.



The first reaction is favored at high temperatures and in the presence of concentrated acid.

The trihydrate is very soluble in water and ethanol. Decomposition begins around 80°C upon formation of the basic salt. At temperatures of 180°C the oxide is produced.

Copper Oxides. Copper(I) oxide [1317-39-1] is a cubic or octahedral naturally occurring mineral known as cuprite [1308-76-5]. It is red or reddish brown in color. Commercially prepared copper(I) oxides vary in color from yellow to orange to red to purple as particle size increases. Usually copper(I) oxide is prepared by pyrometallurgical methods. It is prepared by heating copper powder in air above 1030°C or by blending copper(II) oxide with carbon and heating to 750°C in an inert atmosphere. A particularly air-stable copper(I) oxide is produced when a stoichiometric blend of copper(II) oxide and copper powder are heated to 800–900°C in the absence of oxygen. Lower temperatures can be used if ammonia is added to the gas stream (27–29).

Various hydrometallurgical processes can be used to prepare copper(I) oxide. Acidification with sulfuric acid of ammonia complexes of copper(I) precipitates a red product. Vacuum distillation (30) of Cu₂(NH₃)₄CO₃ produces a stable red copper(I) oxide. Addition of sodium hydroxide to the same carbonate initially gives a yellow material which, on heating, turns orange. A boiling slurry of basic copper sulfate and copper sulfate solution can be reduced using sulfur dioxide to give a red product (31). Solutions of sodium copper(I) chloride can be neutralized using sodium hydroxide to give products of various colors depending on conditions (Table 1). Electrolytic processes can produce copper(I) oxide using copper electrodes in brine. Yellow material is produced at room temperature; orange and red products are made as the temperature is increased.

Table 1. Copper(I) Oxides from the Reaction 2NaCuCl₂ + NaOH → Cu₂O + H₂O + 4NaCl

pH	Temperature, °C	Product		Reference
		Color	Particle size, μm	
7.0		yellow	0.4	32
8.5	60	orange	1	33
alkaline	138	red	2.5	34
10.0 ^a	55	purple	48	32

^a Reactants added simultaneously.

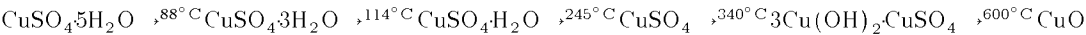
Copper(I) oxide is stable in dry air, but reacts with oxygen to form copper(II) oxide in moist air. Cu₂O is insoluble in water, but dissolves in ammonia or hydrochloric acid. The product disproportionates to copper metal and copper(II) in dilute sulfuric or nitric acid.

Copper(II) oxide [1317-38-0], CuO, is found in nature as the black triclinic tenorite [1317-92-6] or the cubic or tetrahedral paramelaconite [71276-37-4]. Commercially available copper(II) oxide is generally black and dense although a brown material of low bulk density can be prepared by decomposition of the carbonate or hydroxide at around 300°C, or by the hydrolysis of hot copper salt solutions with sodium hydroxide. The black product of commerce is most often prepared by evaporation of Cu(NH₃)₄CO₃ solutions (35) or by precipitation of copper(II) oxide from hot ammonia solutions by addition of sodium hydroxide. An extremely fine (10–20 nm) copper(II) oxide has been prepared for use as a precursor in superconductors (36).

Copper(II) oxide is less often prepared by pyrometallurgical means. Copper metal heated in air to 800°C produces the copper(II) oxide. Decomposition of nitrates, carbonates, and hydroxides at various temperatures also occurs.

Copper(II) oxide is insoluble in water, but readily dissolves in mineral acid or in hot formic or acetic acids. CuO slowly dissolves in ammonia solution, but alkaline ammonium carbonate solubilizes it quickly.

Copper(II) Sulfates. Copper(II) sulfate pentahydrate [7758-99-8], CuSO₄·5H₂O, occurs in nature as the blue triclinic crystalline mineral chalcantite [13817-21-5]. It is the most common commercial compound of copper. The pentahydrate slowly effloresces in low humidity or above 30.6°C. Above 88°C dehydration occurs rapidly.



The solubility and solution density as a function of temperature and acid concentration are given in Figure 1 (37). Crystallization of the pentahydrate can be effected both by concentrated and cooling or by addition of sulfuric acid. The former method is usually preferred because the crystals are of better structural integrity and thus more resistant to hard cake formation. Copper(II) sulfate can be prepared by dissolution of oxides, carbonates, or hydroxides in sulfuric acid solutions. Whereas copper metal does not displace hydrogen from acidic solution, aeration or oxygenation of hot dilute aqueous sulfuric acid in the presence of copper metal is a commonly used commercial method for CuSO₄ preparation (38). Solvent extraction is used to produce the pentahydrate from a variety of copper-containing liquors (4,37).

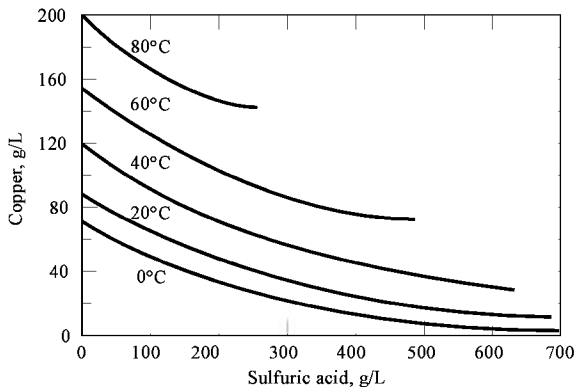
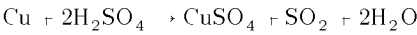


Fig. 1. Aqueous solubility of copper in copper sulfate as a function of sulfuric acid concentration at various temperatures. Reprinted with permission (37).

Copper(II) sulfate monohydrate [10257-54-2], CuSO₄·H₂O, which is almost white in color, is hygroscopic and packaging must contain moisture barriers. This product is produced by dehydration of the pentahydrate at 120–150°C. Trituration of stoichiometric quantities of copper(II) oxide and sulfuric acid can be used to prepare a material of limited purity. The advantages of the monohydrate as opposed to the pentahydrate are lowered freight cost and quickness of solubilization. However, these advantages are offset by the dustiness of the product and probably less than one percent of copper sulfate is used in the monohydrate form.

Anhydrous copper(II) sulfate [7758-98-7] is a gray to white rhombic crystal and occurs in nature as the mineral hydrocyanite. CuSO₄ is hygroscopic. It is produced by careful dehydration of the pentahydrate at 250°C. An impure product can also be produced from copper metal and hot sulfuric acid:



There are four basic sulfates that can be identified by potentiometric titration using sodium carbonate (39,40): langite [1318-78-1], CuSO₄·3Cu(OH)₂·H₂O; brochantite [12068-81-4], CuSO₄·3Cu(OH)₂; antedite [12019-54-4], CuSO₄·2Cu(OH)₂; and CuSO₄·CuO·2Cu(OH)₂·xH₂O. The basic copper(II) sulfate that is available commercially is known as the tribasic copper sulfate [12068-81-4], CuSO₄·3Cu(OH)₂, which occurs as the green monoclinic mineral brochantite. This material is essentially insoluble in water, but dissolves readily in cold dilute mineral acids, warm acetic acid, and ammonia solutions.

Tribasic coppersulfate is usually prepared by reaction of sodium carbonate and copper sulfate. As the temperature of the reaction contents increases so does the size of the resulting particle. For use as a crop fungicide, intermediate (40–60°C) temperatures are used to obtain a fine particle. When lower temperatures are used to precipitate basic copper(II) sulfate, products high in sulfate and water of hydration are obtained.

Economic Aspects

It has been estimated that 100,000 t of copper is used in the production of copper compounds annually (41) representing less than 1% of the total production of primary metal (42). As much as 70,000 t yr (as copper) is used in agriculture for fungicides, animal feeds, and crop nutrients (43). Table 2 gives estimates for agricultural use by region of the world. Algicides make up the second largest group of copper compounds, followed by wood (qv) treatment and antifouling pigments (qv), mining flotation (qv), electroplating (qv), colorants, electronics, and catalysts.

Table 2. Annual Agriculture Usage of Copper^a

Region	Total, t	Usage		
		Fungicide, %	Animal feed, %	Crop nutrient, %
Europe	35,293	68.4	21.5	10.1
Asia Australia	8,907	41.8	1.5	56.6
Africa	7,385	98.7	0.2	1.1
America	18,750	6.8	4.3	25.9
Total	70,335	76.4	12.1	11.4

^a Ref. 43.

The price of the compounds of copper varies with the price of the metal and the grade of the product. Average Commodities Exchange (COMEX) copper metal prices per pound from 1970 to November of 1992 are as follows:

Year	Price
1970	\$0.620
1975	\$0.553
1980	\$0.961

Year	Price
1985	\$0.610
1990	\$1.183
1991	\$1.049
1992	\$1.030

Analysis

The specifications or typical analyses of selected copper compounds are given in Table 3 (44,45).

Table 3. Analyses of Copper Compounds

Assay	Composition, wt %					
	Cu(OH) ₂ ·CuCO ₃		Cu(NO ₃) ₂ ·3H ₂ O	CuO	CuSO ₄ ·5H ₂ O	CuSO ₄ ·5H ₂ O
grade	light	solution	black	technical	electronic	
Cu	55.4	14.0	78.5	25.3	25.5	
Fe	0.1	0.002	0.01	0.01	0.003	
Pb	0.005	0.002	0.01	0.005	0.001	
Zn	0.015	0.0005	0.05	0.01	0.0005	
insolubles	0.05 ^a		0.02 ^a	0.05 ^b	0.008 ^b	
ABD ^c , kg m ³	300	1450	1300	1000	1000	

^a Insoluble in HNO₃.
^b Insoluble in H₂O.
^c Apparent or free-falling bulk density.

Separation. Preliminary separation of copper from a variety of interfering elements by extraction using diethyldithiocarbamate or dithizone has been practiced since the early 1940s (46–52) and is still commonly used (53,54). Excellent separations from a complex matrix can be effected using 2-hydroxyoximes (55), and precipitation of copper as the sulfide is commonplace. Hydrolysis to give basic salts of copper is quite often used, using collectors such as lanthanum (56) or iron(III) (57,58). Numerous ammonia-insoluble impurities in copper can be separated by coprecipitation with lanthanum using ammonia (59). Selective precipitation with trithiocarbonate ion (60), or concentration using ammonium pyrrolidine dithiocarbamate [5108-96-3] (APDC) or dithiooxamide (61) can be used. Ion exchange (qv) using Chelex 100 has been used to preconcentrate trace copper (62–65) Chelating ion exchangers have been developed for selective preconcentration (66) or are used in hydrometallurgical applications (67) (see CHELATING AGENTS; METALLURGY; TRACE AND RESIDUE ANALYSIS).

Determination. The most accurate (68) method for the determination of copper in its compounds is by electrogravimetry from a sulfuric and nitric acid solution (45). Pure copper compounds can be readily titrated using ethylene diamine tetracetic acid (EDTA) to a SNAZOXs or Murexide endpoint. Iodometric titration using sodium thiosulfate to a starch–iodide endpoint is one of the most common methods used industrially. This latter titration is quicker than electrolysis, almost as accurate, and much more tolerant of impurities than is the titration with EDTA. Gravimetry as the thiocyanate has also been used (68).

Colorimetric procedures are often used to determine copper in trace amounts. Extraction of copper using diethyldithiocarbamate can be quite selective (60,62), but the method using dithizone is preferred because of its greater sensitivity and selectivity (50–52). Atomic absorption spectroscopy, atomic emission spectroscopy, x-ray fluorescence, and polarography are specific and sensitive methods for the determination of trace level copper.

Health and Safety

Copper is one of the twenty-seven elements known to be essential to humans (69–72) (see MINERAL NUTRIENTS). The daily recommended requirement for humans is 2.5–5.0 mg (73). Copper is probably second only to iron as an oxidation catalyst and oxygen carrier in humans (74). It is present in many proteins, such as hemocyanin [9013-32-3], galactose oxidase [9028-79-9], ceruloplasmin [9031-37-2], dopamine β-hydroxylase, monoamine oxidase [9001-66-5], superoxide dismutase [9054-89-1], and phenolase (75,76). Copper aids in photosynthesis and other oxidative processes in plants. Copper is toxic in exceedingly low concentrations to most fungi, algae, and certain bacteria and can be lethal to higher life forms in relatively high doses. The LD₅₀'s in mg kg for rats are copper(II) acetate, 710; copper(II) chloride dihydrate, 163; anhydrous copper(II) nitrate, 940; and copper(II) sulfate pentahydrate, 300. For female rats, copper hydroxide has an LD₅₀ of 2200 mg kg. The acute oral toxicity in humans, LD_{LO}, is 100 mg kg, but recovery from ingestion of 600 mg kg has occurred. The symptoms of copper poisoning are nausea, vomiting, cramps, gastric disturbances, apathy, anemia, convulsions, coma, and death. Chronic toxicity from copper poisoning has not been definitely observed although there are numerous accounts of chronic poisoning resulting from the refining of copper. This poisoning is now thought to be a result of the common impurities in copper ores such as lead, selenium, and arsenic (77,78). Copper is known to accumulate in the liver, but it is uncertain whether pathological changes occur. Inhalation of dusts can cause metal fume fever (79,80), and ulceration or perforation of the nasal septum. Mild discomfort has been noted with workplace concentrations as low as 0.08 mg m³. The workplace standard (TLV) for copper dusts or mist is 1 mg m³ and 0.2 mg m³ for copper fume (81).

Uses

Examples of uses (82) of copper compounds are given in Table 4 which lists the materials of primary industrial importance. The majority of copper compounds are used as fungicides, nutritionals, and algicides.

Table 4. Uses of Copper Compounds

Compound	CAS Registry Number	Molecular formula	Uses
copper(I) acetate	[598-54-9]	CuCH ₃ COO	absorption of olefins
copper(II) acetate	[142-71-2]	Cu(CH ₃ COO) ₂	fabrics, textiles, pigment, catalyst
copper(II) acetate monohydrate	[6046-93-1]	Cu(CH ₃ COO) ₂ ·H ₂ O	
copper(II) acetate, basic	[52503-63-6]	Cu(CH ₃ COO) ₂ ·CuO·6H ₂ O	manufacture of Paris Green fungicides, pigments, textiles
copper(II) arsenate	[7778-41-8]	Cu ₃ (AsO ₄)·4H ₂ O	insecticides, wood preserving, antifouling pigment
copper(I) bromide	[7787-70-4]	CuBr	catalyst
copper(II) bromide	[7789-45-9]	CuBr ₂	catalyst, brominating reagent, intensifier (photography)
copper(II) carbonate, basic	[12069-69-1]	CuCO ₃ ·Cu(OH) ₂	fungicides, animal feeds, catalyst, oil treatment
copper(I) chloride	[7758-89-6]	CuCl	catalyst, absorption of CO, fuel oil treatment
copper(II) chloride	[7447-39-4]	CuCl ₂	catalyst, mordant, electroplating, pigment
copper(II) chloride dihydrate	[10125-13-0]	CuCl ₂ ·2H ₂ O	

copper(II) chloride hydroxide	[1332-65-6]	$\text{CuCl}_2 \cdot 3\text{Cu}(\text{OH})_2$	fungicides, pigment
copper(II) chromate(VI)	[13548-42-0]	CuCrO_4	wood preserving, textiles
copper(II) chromate(III)	[12018-10-9]	CuCr_2O_4	hydrogenation catalyst
copper(I) cyanide	[544-92-3]	CuCN	electroplating, catalyst
copper(II) formate	[544-19-4]	$\text{Cu}(\text{HCO})_2$	mildewcide, bactericide
copper(II) fluoborate	[38465-60-0]	$\text{Cu}(\text{BF}_4)_2$	electroplating, electronics
copper(II) gluconate	[527-09-3]	$\text{Cu}(\text{C}_6\text{H}_{11}\text{O}_7)_2$	dietary supplement, breath freshener
copper(II) hydroxide	[20427-59-2]	$\text{Cu}(\text{OH})_2$	fungicides, rayon manufacture, catalyst, antifouling pigment, electrolysis, electroplating
copper(I) iodide	[7681-65-4]	CuI	heat and light stabilizer in polymers, photographic emulsions, and light-sensitive paper, and in oil drilling to aid in corrosion inhibition in highly acid environments; used as feed additive, in cloud seeding, and as a double salt with mercury(II) iodide as a temperature indicator
copper(II) naphthenate	[1338-02-9]		fungicide, mildewcide in textiles, woods, and paints
copper(II) nitrate	[3251-23-8]	$\text{Cu}(\text{NO}_3)_2$	electrolysis and electroplating, electronics, fuel oil treatment, colorant, pyrotechnics, catalyst
Copper(II) nitrate trihydrate	[10031-43-3]	$\text{Cu}(\text{NO}_3)_2 \cdot 3\text{H}_2\text{O}$	
copper(II) oleate	[1120-44-1]	$\text{C}_{18}\text{H}_{34}\text{O}_2 \cdot 1/2\text{Cu}$	fuel oil combustion improver, emulsifier, dispersant, antifouling coating for fish nets and lines
copper(II) oxalate	[814-91-5]	CuC_2O_4	catalyst, stabilizer
copper(I) oxide	[1317-39-1]	Cu_2O	antifouling pigment, fungicide, pigment, catalyst
copper(II) oxide	[1317-38-0]	CuO	wood preserving, feed additive, pigment, catalyst
copper(II) phosphate trihydrate	[7798-23-4]	$\text{Cu}_3(\text{PO}_4)_2 \cdot 3\text{H}_2\text{O}$	fungicide, corrosion inhibitor
copper(II) diphosphate hydrate	[10102-90-6]	$\text{Cu}_2\text{P}_2\text{O}_7 \cdot x\text{H}_2\text{O}$	electroplating plastics, aluminum, and zinc
copper(II) stearate	[660-60-6]	$\text{C}_{18}\text{H}_{33}\text{O}_2 \cdot 1/2\text{Cu}$	antifouling paints, wood and textile preservation
copper(II) sulfate	[7758-98-7]	CuSO_4	fungicides, algicides, antifouling paints, electrolysis and electroplating, electronics, fuel oil treatment, wood preserving
copper(II) sulfate pentahydrate	[7758-99-8]	$\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$	
copper(II) sulfate, tribasic	[12068-81-4]	$\text{CuSo}_4 \cdot 3\text{Cu}(\text{OH})_2$	fungicides
copper(I) sulfide	[22205-45-4]	Cu_2S	luminous paints, solar cells, semiconductors, lubricants
copper(II) sulfide	[1317-40-4]	CuS	antifouling pigment, manufacture of aniline black
copper(I) thiocyanate	[1111-67-7]	CuSCN	antifouling pigment

Foliar Fungicides and Bactericides. Of the ~70,000 t yr as copper in compounds used in agriculture, almost 75% is used in the control of fungi (see FUNGICIDES, AGRICULTURAL). The first reference to the use of copper as a fungicide dates to 1761 (83) where copper sulfate was used on wheat seed for the control of bunt. In 1807 (84) the discovery of copper as a fungicide was made and the discovery of Bourdeaux mixture (copper sulfate plus lime) followed in 1882.

In order for copper to have any persistence as a foliar fungicide it must be rendered insoluble so that it is not washed off the leaf surface during rainfall. Some of the most common copper fungicides are given in Table 5 (85,86). Copper compounds have been used successfully since the latter 1880s on fruit, vegetables, nut crops, and ornamentals. Copper is an effective broad-spectrum fungicide, although its action is more prophylactic in nature.

Table 5. Copper Fungicides

Active ingredient	Year introduced	Examples	Manufacturer	Copper, %
copper sulfate	1761			25
copper sulfate + lime	1873	Bourdeaux Mixture		<20
copper sulfate + soda ash	>1885	Burgundy Mixture		<20
copper carbonate	1887	Basic Copper Carbonate	Agtrol	55
		C.C.D.	Duclos	12.5
copper oxychloride	1900	Cupravit	Bayer	50
		Vitigran	Hoechst	50
		Cobox	BASF	50
		Microcop	Tennessee Chem.	50
		Coprantol	CIBA-GEIGY	30
		COC 50WP	Agtrol	50
		Cuprossil	Scam	50
		Blaukupfer	Sandoz	50

copper ammonia complex	1917	Copper-Count-N	Mineral Research	8
		K-Cop	Griffin	8
		ALGAE-RHAP CU-7	Agtrol	7
		Copac-E	BASF	28–30 ^a
basic copper sulfate	1930	Basic Copper Sulfate	Agtrol	
		Tribasic Copper Sulfate	Tennessee Chem.	
copper(I) oxide	1932	Cobre Nordox	Quimica Masso	50
		Cobre Sandoz	Sandoz	50
		Oxiram 50 PM	CIBA-GEIGY	50
copper hydroxide	1960	Kocide 101	Griffin	50
		Kocide DF	Griffin	40
		Champion	Agtrol	50
		CHAMP	Agtrol	15
		Blue-Shield	Cuproquim	50

^a Value given is g L.

Whereas copper is not as effective as some organics against specific pathogens, pathogen resistance to copper has been minimal in the >100 year usage. Many organic fungicides are used less often then copper compounds because of residue tolerance, grower reluctance, or for regulatory reasons. Use of copper compounds as fungicides has seen a resurgence in the latter part of the twentieth century.

Plant and Animal Nutrient. Copper is one of seven micronutrients that has been identified as essential to the proper growth of plants (87). Cereal crops are by far the most affected by copper deficiency (see WHEAT AND OTHER CEREAL GRAINS). Greenhouse studies have shown yield increases from 38° o to over 500° o for wheat, barley, and oats (88) using copper supplementation. A tenfold increase in the yield of oats was reported in France (89). Symptoms of copper deficiency vary depending on species, but often it is accompanied by withering or chlorosis in the leaves that is not ammenable to iron supplementation. In high concentrations, particularly in low pH soils, copper can be toxic to plants.

Copper compounds are used as feed additives in Europe and the United States primarily for chickens and swine (see FEEDS AND FEED ADDITIVES) (90,91). Copper increases the rate of gain and feed efficiencies of the animals. It is unclear whether this results from overcoming animal deficiencies or by enhancing preservation of feedstuffs.

Algicides. Copper sulfate has been used to control algae in lakes and reservoirs since around 1905. In the United States it is used by every state either on a routine or an occasional basis (92). It provides an effective and ecologically sound method to control algal blooms. Algae cause water supplies to have odor and taste. The turbidity and scum that forms from algal blooms can blind filters and affect certain industries adversely. Significant algal blooms decrease recreational use and are a primary cause of oxygen depletion that results in fish kills.

Copper sulfate is by far the most common algicide. Other copper-containing algicides for use in domestic applications such as swimming pools are usually chelated to prevent hydrolysis and precipitation of the copper.

Other Uses. Copper is one of the primary ingredients used in wood preservation. In combination with chromium and arsenic or zinc or borates it has largely replaced pentachlorophenol and creosote. As an antifouling pigment, copper(I) oxide has found renewed use because of regulatory problems of organotin compounds (see COATINGS, MARINE). Copper is much more amenable to the environment than the tin or lead alternatives. Copper phthalocyanines are excellent color-stable blue and green pigments for paints (see PHTHALOCYANINE COMPOUNDS). As an oxidation catalyst, copper with chrome is used extensively. Mixed oxides of chromium and copper show promise as a replacement for platinum group metals in emission control devices. Copper used with zinc or with zinc and chromium is employed as a low temperature shift catalyst (LTS) for the synthesis of fuels and methanol from coal gases (see COAL CONVERSION PROCESSES; FUELS, SYNTHETIC). Electroplating uses copper sulfate baths extensively and electroless copper baths are used to produce circuit boards (see ELECTROLESS PLATING). Copper oxide is used in air-bag technology, and oxides and sulfides are used as solar collectors (see SOLAR ENERGY). Copper compounds are also used as fuel additives to minimize sulfide and carbon monoxide emissions.

BIBLIOGRAPHY

"Copper Compounds" in *ECT* 1st ed., Vol. 4, pp. 467–479, by L. G. Utter and S. B. Tuwiner, Phelps Dodge Corp.; in *ECT* 2nd ed, Vol. 6, pp. 265–280, by E. A. Winter, M. J. Montesinas, and J. E. Singley, in *ECT* 3rd ed., Vol. 7, pp. 97–109, by R. N. Kust, Kennecott Copper Corp.

1. W. Hatfield and R. Whyman, in R. Carlin, ed., *Transition Metal Chemistry*, Vol. 5, Marcel Dekker, New York, 1969, pp. 47–179.
2. C. P. Poole, Jr., T. Datta, and H. A. Farach, with M. M. Rigney and C. R. Sanders, *Copper Oxide Superconductors*, John Wiley & Sons, Inc., New York, 1988.
3. W. E. Hatfield and J. H. Miller, Jr., eds., *High-Temperature Superconducting Materials, Preparations, Properties, and Processing*, Marcel Dekker, Inc., New York and Basel, 1988. Original paper: J. G. Bedmorz and K. A. Muller, *Z. Phys. B-Condensed Matter* **64**, 189 (1986).
4. C. Merigold, D. Agers, and J. House, *International Solvent Extraction Conference*, London, 1971.
5. **Ger. Pat. 1000361** (1955), F. Bittner (to Degussa).
6. **Fr. Pat. 2009852** (1969), (to Degussa).
7. **Ger. Pat. 1813891** (1958), (to Norddeutsche Affinerie).
8. **U.S. Pat. 3,679,359** (1972), E. Haberland and W. Perkow (to Norddeutsche Affinerie).
9. R. Keller and H. Wycoff, *Inorg. Synth.* **2**, 1–4 (1946).
10. **Ger. Pat. 3305545** (1983), E. Mack and L. Witzke (to Goldschmidt).
11. **Ger. Pat. 1080088** (1958), H. Niemann and K. Herrmann (to Schering Corp.).
12. **U.S. Pat. 2,367,153** (1945), C. Swinehart (to Harshaw).
13. **U.S. Pat. 3,202,478** (1965), A. Hindel, S. Raval, S. Damani, H. Damani, and K. Damani (to Sudhir Chem. Co.); **Brit. Pat. 912,125** (1962).
14. **Ger. Pat. 1159914** (1963), E. Podschus (to Bayer AG).
15. **PL 55953** (1968), (to Instytut Przemyslu Org.).
16. **U.S. Pat. 1,800,828** (1931), W. Furness (to Cellocilk Co.).
17. **U.S. Pat. 2,104,754** (1938), D. Marsh and B. Marsh.
18. **U.S. Pat. 2,536,096** (1951), P. Rowe (to Mountain Copper).
19. **U.S. Pat. 25,225,242** (1950) P. Rowe (to Lake Chem. Co.).
20. **U.S. Pat. 24,324** (1957), W. Furness (to Copper Research).
21. **Ger. Pat. 1592441** (1965), W. Furness (to Kennecott Copper); **U.S. Pat. 3,194,749** (1965), (to Kennecott Copper).
22. **U.S. Pat. 3,231,464** (1966), E. Dettwiler and J. Fillietaz (to Rohm and Haas).
23. **U.S. Pat. 4,418,056** (1983), M. Gonzalez (to Cuproquim S.A.).
24. **Brit. Pat. 83/01912** (1983), J. Giuliani and A. Meyer (to Giuliani Adolfomer Ind.).
25. **U.S. Pat. 4,490,337** (1984), H. W. Richardson (to Kocide Chemical Corp.).
26. **U.S. Pat. 4,808,406** (1989), N. C. Brinkman (to Kocide Chemical Corp.).
27. **U.S. Pat. 3,466,143** (1969), H. Day (to Calumet and Hecla).
28. **U.S. Pat. 2,758,014** (1956), J. Drapeau and P. Johnson (to Glidden).
29. **U.S. Pat. 2,891,842** (1959), J. Drapeau and P. Johnson (to Glidden).

30. U.S. Pat. **3,492,115** (1970) (to S. Mahalla).

31. U.S. Pat. **2,665,192** (1954) P. Rowe (to Mountain Copper).

32. **Jpn. Kokai 78/133775** (1978), (to Nippon Chem.).

33. **Brit. Pat. 936,922** (1963), A. Campbell and A. Taylor (to ICI).

34. **Jpn. Kokai 80/71629** (1980), (to Nippon Chem.).

35. W. Kunda, H. Veltman, and D. Evans, *Copper Met. Proc. Extr. Met. Div. Symp.*, 27–69 (1970).

36. "Dowa Mining," *Adv. Ceram. Rep. Ang.* (1990).

37. H. Moyer, *AIIME Annual Meeting*, New Orleans, La., Feb. 18–22, 1979.

38. U.S. Pat. **2,533,245** (1950), G. Harike (to Tennessee Copper).

39. L. Markov and K. Balarev, *Izv. Khim* **15**, 472–481 (1982).

40. H. Weiser, W. Milligan, and E. Cook, *J. Am. Chem. Soc.* **64**, 503–508 (1942).

41. Commodities Research Unit, *Copper Stud.* **15**, 5–10 (1988).

42. J. L. W. Jolly and D. L. Edelstein, *Copper, Annual Report 1990*, Dept. of the Interior, U.S. Bureau of Mines, Washington, D.C., Apr. 1992.

43. V. Shorrocks, *Use of Copper as a Micronutrient for Crops, Part I*, Planning Study, INCRA, New York, July 1984.

44. Z. Marczenko, *Separation and Spectrophotometric Determination of Elements*, Ellis Horwood Ltd., West Sussex, UK, 1986; E. B. Sandell, *Colorimetric Determination of Traces of Metals*, Interscience Publishers, New York, 1959.

45. C. Freedenthal, *Copper Compounds, Encyclopedia of Industrial Chemical Analysis*, John Wiley & Sons, Inc., New York, 1970, pp. 651–680.

46. L. I. Butler and H. O. Allen, *J. Assoc. Office. Agr. Chem.* **25**, 567 (1942).

47. L. Gerber, R. I. Claassen, and C. S. Boruff, *Ind. Eng. Chem., Anal. Ed.* **14**, 364 (1942).

48. I. Stone, *Ind. Eng. Chem., Anal. Ed.* **14**, 479 (1942).

49. T. C. J. Ovenston and C. A. Parker, *Anal. Chim. Acta* **4**, 135 (1950).

50. G. H. Bendix and D. Grabenstetter, *Ind. Eng. Chem., Anal. Ed.* **15**, 649 (1943).

51. S. L. Morrisson and H. L. Paige, *Ind. Eng. Chem., Anal. Ed.* **18**, 211 (1946).

52. P. M. Heertjes, *Chem. Weekblad* **42**, 91 (1946).

53. P. L. Schuller and L. E. Coles, *Pure Appl. Chem.* **51**, 385 (1979).

54. E. A. Allen, P. K. N. Bartlett, and G. Ingram, *Analyst* **109**, 1075 (1984).

55. R. Dubczynski and H. Maleszewska, *Chem. Anal.* **32**, 619 (1987).

56. Z. Marczenko and K. Kasiura, *Chem. Anal. (Warsaw)* **10**, 449 (1965).

57. N. A. Rudnev, G. I. Malofeeva, N. P. Andreeva, and T. V. Tikhonova, *Zh. Analit. Khim.* **26**, 697 (1971).

58. A. I. Novikov, A. A. Shaffert, and E. K. Schekoturova, *Zh. Analit. Khim.* **32**, 1108 (1977).

59. ASTM, *Chemical Analysis of Metals and Metal Bearing Ores*, VO3.05, P195, Philadelphia, Pa., 1987.

60. K. Singh, R. Gupta, and P. Bhatia, *Acta Chim. Hung.* **113**, 3 (1983).

61. R. Santelli, M. Gallego, and M. Valcarcel, *Anal. Chem.* **61**, 1427 (1989).

62. W. Van Berkel, A. Overbosch, G. Feenstra, and F. Maessen, *J. Anal. At. Spectrom.* **3**, 249 (1988).

63. S. Pai, T. Chen, and G. Wong, *Anal. Chem.* **62**, 774 (1990).

64. J. P. Riley and D. Taylor, *Anal. Chim. Acta* **40**, 479 (1968).

65. H. Watanabe and H. Ohmori, *Talanta* **28**, 774 (1981).

66. A. Warshawsky, in M. Streat and D. Naden, eds., *Ion Exchange and Sorption Processes in Hydrometallurgy*, Society of Chem Industry, Chichester, UK, 1987, p. 166.

67. J. Melling and D. West, in D. Naden and M. Streat, eds., *Ion Exchange Technology*, Ellis Horwood Ltd., UK, p. 724.

68. I. Sarudi, *Z. Anal. Chem.* **130**, 301 (1950).

69. H. E. Stokinger, *Patty* **2A**, 1620–1630 (1981).

70. S. Cohen, *JOM, J. Occup. Med.* **16**, 621–624 (1974).

71. W. Dreichmann and H. Gerarde: *Toxicology of Drugs and Chemicals*, Academic Press, Inc., New York, 1969.

72. W. Mertz, *Nutr. Today* **18**, 26 (1983).

73. *A Critical Review of Copper in Medicine Report No. 234*, International Copper Research Association, Inc., Goteborg-New York, 1975.

74. D. Kertesz, R. Zito, and F. Ghiretti, in O. Hayaishi, ed., *Oxygenases*, Academic Press, Inc., London, 1962.

75. *Biological Roles of Copper*, no. 79, Ciba Foundation Symposium, Excerpta Medica and Elsevier-North Holland, Amsterdam, The Netherlands, 1980, p. 343.

76. H. Siegel, ed., *Metal Ions in Biological Systems*, Vol. 12, *Properties of Copper*, Marcel Dekker, New York, 1981, p. 384.

77. E. Browning, *Toxicity of Industrial Metals*, 2nd ed., Butterworths, London, 1969.

78. R. Fabre and R. Truhaut, *Precis de Toxicologie*, Centre Doc. University, Paris, 1960.

79. A. Askergren and M. Mellgren, *Scand. J. Work, Environ. Health*, **1**, 45–49 (1975).

80. R. Gleason, *Am. Ind. Hyg. Assoc. J.* **29**, 375–376 (1968).

81. *American Conference of Governmental Industrial Hygienists (ACGIH)*, Cincinnati, Ohio, 1991, p. 121.

82. H. W. Richardson, ed., *Copper Compounds Application Handbook*, Marcel Dekker, New York, 1992.

83. H. Shulthess, *Abhandl. Naturf. Gesell. Zurich*, **1**, 498 (1761).

84. *Prevost, Phytopath. Classic* **6**, 1 (1807).

85. W. T. Thomson, *Agricultural Chem. Book IV, Fungicides*, Thomson Publications, Fresno, Calif., 1991.

86. H. Kidd, D. Hadley, and J. M. Kennedy, eds., *European Directory of Agrochemical Products*, Vol. 1, *Fungicides*, 2nd ed., The Royal Society of Chemistry, UK, 1986.

87. U. Gupta, "Copper in Agricultural Crops," in J. O. Nriagu, ed., *Copper in the Environment*, Part 1, John Wiley & Sons, Inc., New York, 1979.

88. U. Gupta and L. B. McLeod, *Can. J. Soil. Sci.* **50**, 373–378 (1970).

89. L. Duval, *C. R. Acad. Agric. Fr.* **49**, 1216–1220 (1963).

90. G. Cromwell, T. Stahly, and W. Williams, *Feedstuffs* **53**, 30, 32, 35, 36 (1982).

91. L. DeGoey, R. Wahlstrom, and R. Emerick, *J. Anim. Sci.* **33**, 52–57 (1971).

92. D. M. McKnight, S. W. Chrisholm, F. M. M. Morel, INCRA Project No. 252, *Copper Sulfate Treatment of Lakes and Reservoirs: Chemical and Biological Considerations*, Ralph M. Parsons Lab, Division of Water Resources & Hydrodynamics, Massachusetts Institute of Technology, Cambridge, Mass., 1981.

H. Wayne Richardson
CP Chemicals, Inc.

COPYRIGHTS AND TRADEMARKS

Copyrights,
Trademarks,

COPYRIGHTS

Copyright has been grouped with other forms of legal protection under the general term *intellectual property*. It is a means of protecting that particular form of creativity which has been variously referred to as originality of authorship, expression of ideas, or writings of an author. It is distinct from forms of intellectual property that do not protect original expression of authorship, eg, patents, which protect novel inventions or discoveries; trademarks, which protect terms identifying the source of origin of goods and services; and trade secrets which protect confidential, proprietary information.

Historical Background.

There is considerable scholarly speculation on ancient precursors of copyright. However, no significant copyright protection existed in theory or practice until the development of the printing press, a means of mass duplication of creative work. The law did not need a coherent protective structure for written and graphic works when the only means of copying them, by hand, was so labor-intensive as to require an investment potentially far in excess of the worth of the copy.

The invention of the printing press meant that written and graphic works could be duplicated and disseminated in vast numbers. Hence, the necessity of protecting the right to copy, ie, the copy right, arose. Copyright protection in England emerged in the fifteenth and sixteenth centuries as a means of censorship and control, rather than as a means of encouraging authorship. The Crown feared the dissemination of treasonous writings; the Church feared the spread of heretical writings, and so no work could be published other than through the Stationers' Company, ie, the printers' trade guild, which was under the control of the Crown. Those who were allowed to publish through the Stationers' Company gained copyright protection for their works.

In the late seventeenth century, the exclusive grant of copyright through the Stationers' Company expired. In 1710, the English Parliament enacted the first true copyright statute, the Statute of Anne, which granted copyright protection to the author, as creator of the work, rather than to the printer, as exploiter of the work. This statute also limited the duration of copyright protection to a fixed term of years, after which the work would go into the public domain and be free for all to use.

In the United States, immediately after independence, 12 of the 13 original states enacted their own copyright statutes, in part as a result of lobbying and efforts by Noah Webster and James Madison. However, because copyright easily transcends state borders, and because of the lack of effective enforcement and the inhospitability of state courts to out-of-state copyright owners, the need for a uniform, national copyright law became apparent. Thus the Constitutional Convention included the power to enact a national copyright, as well as patent, law among the enumerated powers of Congress (1).

The purpose of copyright was to "promote progress" in learning (the eighteenth century sense of the word *science*) for the good of all society. Economic property rights were granted to creators as incentive to pursue their vocation and so enrich society. But there were limitations on those property rights: they could only endure for "limited times," and could only be granted for "writings" of "authors."

The First Congress after the Constitution was adopted enacted a national copyright statute in 1790. Thereafter, the development of federal copyright law followed a pattern that endures to this day. As developments in technology and forms of mass entertainment, communications, scholarship, and the arts have occurred, the copyright law has been amended to accommodate them. Periodically, a total revision of the copyright statute has been necessary. Such complete revisions occurred in 1831, 1870, 1909, and 1976.

The 1976 Copyright Act is the basic law governing copyright in the late twentieth century (2). But the principles and provisions of the 1909 act are still important, eg, in the provisions regarding duration. There have been many significant amendments to the 1976 act since it went into effect on January 1, 1978.

Copyrightability

Under U.S. law, a work is either protected, ie, copyrighted, or unprotected and free for all to use, ie, in the public domain. Once a work enters the public domain, it cannot thereafter be recovered and protected again.

The Copyright Act specifies that copyright extends to "original works of authorship fixed in any tangible medium of expression, now known or later developed, from which they can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device" (3). Many of the requirements for copyrightability may be gleaned from this provision.

The copyrightable work must be an original work of authorship. Thus originality, and not novelty as for patent rights, constitutes the touchstone of protection. Courts have defined what makes a work original both positively, ie, the spark of creativity, or something ... which is one person's alone; and negatively, ie, as constituting that which has not been copied from another, even if not unique or novel (4,5).

The work must also be the product of an author. That is, a human being has created the work, even if at the behest or for the ownership of a corporate entity and even if a machine or device, such as a camera or a computer program, was used as a tool in the creative process.

Finally, the work must be fixed in a tangible medium of expression. To a very limited extent, there are some works that are not so fixed, such as purely improvised and unrecorded pieces of music or choreography; extemporaneous speeches; or live, unrecorded, and ephemeral broadcasts. Unfixed works are protected by state common law copyright, and not the federal statute. Under the federal law, however, all works that are so fixed are governed exclusively by the federal statute.

The Subject Matter of Copyright. The Copyright Act specifies, by way of example, the types of works that are covered, ie, literary works; musical works, including lyrics; dramatic works, including accompanying music; pantomimes and choreographic works; pictorial, graphic, and sculptural works; motion pictures and other audiovisual works; sound recordings; and architectural works. This list is nonexhaustive (3).

The test for copyrightability is not subjective. It does not matter whether the work is good or bad art, or even obscene. Such considerations are not relevant to copyrightability.

The law also specifies that works of the U.S. government are not subject to copyright protection (6). The United States may own copyrights, but works created by government employees are common property and thus are not copyrightable. However, works created pursuant to government grants or using government funds may be copyrightable and owned by those outside the government who received the funds, depending on the regulations of the particular government agency making the grant. Indeed, if both parties agree, the copyright in such works may be transferred to, and owned by, the government.

The Idea/Expression Dichotomy. Copyright protects the expression of ideas, but not ideas themselves. Thus copyright will not protect any procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied (7).

Copyright does not extend to titles, phrases, or, as a general rule, forms. Such items do not have the requisite originality of expression to distinguish them from the ideas they represent.

Copyright also does not extend to facts or news, but only to the particular form of expression of those facts. For example, the fact that a particular chemical causes a particular reaction may not be protected by copyright. But the text of an article describing that reaction and describing the experimental

procedure to elicit it may be protectable, because those textual descriptions constitute the expression of the fact and not the fact itself. Thus copyright protection does not prevent anyone from recreating the experiment, but only from copying the article that describes it.

Similarly, copyright does not protect research. According to the Supreme Court, copyright protects creativity, not effort, no matter how significant that effort is (4).

In some circumstances the idea and expression are not distinguishable. A jewelry pin made in the form of a bee, or simple sweepstakes rules, may be examples of such a merger of idea and expression. There is no other way, or only a limited number of ways, to depict a bee in gold, or to express those contest rules. In such cases, copyright will not protect the work, for that would protect not only the expression, but the idea itself.

Utilitarian Works. These works may be copyrightable, but only to the extent that they contain copyrightable subject matter (8). That copyrightable subject matter must be physically or conceptually separable from the purely utilitarian object. For example, a common straight-backed chair is not copyrightable, because there is nothing about it that is physically or conceptually separable from its chairness, its purely utilitarian function. However, if that chair contains a carved lion's head on its back, the lion's head is physically or conceptually separable and, therefore, copyrightable.

Compilations. Copyright extends not only to works that can exist on their own but also to compilations of such works or even of public domain material. The Copyright Law imposes a three-step test for such copyrightable compilations. They must first constitute the collection and assembling of preexisting data or materials. Second, those materials must be selected, coordinated, or arranged in a particular fashion. Third, that selection, coordination, or arrangement must itself possess sufficient originality and creativity to constitute an original work of authorship. For example, the alphabetical listing of all subscribers to a telephone company's service, as in an ordinary white-pages telephone directory, does not constitute a copyrightable compilation, that is, no selection was made (all subscribers were listed) and no arrangement or coordination rose to the level of original expression, as the listings were merely ordered alphabetically. On the other hand, the anthologizing of articles on a particular subject, such as in an encyclopedia, does constitute a copyrightable compilation. Compilations of materials that can stand on their own as copyrightable works, such as encyclopedias, journals, or newspapers, are called collective works.

Copyright in a compilation does not extend to, affect, or enlarge the protection of the underlying preexisting materials. Rather, it is only the original expression contributed by the author of the compilation, such as the selection of articles in an encyclopedia, to which the compilation copyright extends.

Copyright Ownership

Copyright is a property right. Although it differs from most other forms of property in that it is intangible, it nevertheless has the essential elements of property and is governed by the principles of property ownership.

The intangible nature of copyright requires a distinction between the intangible property of the copyright, called a work, and the material object in which the copyrighted work is embodied, ie, a copy or phonorecord, which includes such diverse media as paper and ink, computer disks, and audiotapes. Ownership of the copyrighted work does not constitute ownership of the material object in which it is embodied, and vice versa. Copyright ownership vests initially in the author or authors of the work.

Joint Authorship. When more than one author has created a work, the work is said to be a joint work. Under the law, such a joint work is one prepared by two or more authors with the intention that their contributions be merged into inseparable or interdependent parts of a unitary whole. Thus joint authorship can occur when a composer and a lyricist collaborate on a song. Even though their contributions can exist independently of each other, ie, the music as an instrumental, the lyrics as a poem, they were created as interdependent parts of a unitary whole, ie, the song. Two scientists collaborating on an article for a scientific journal are similarly joint authors. Their contributions cannot be teased out of the article they have written and so are inseparable parts of a unitary whole. The test of joint authorship is intention. The creation must be made with the intention that the contributions be merged into one work, and that intention cannot be imputed after the fact of creation if it was not there to begin with.

Joint Ownership. Joint ownership of copyright occurs when there is joint authorship, but it may also occur in other ways, for example, by transfer of a copyright to two or more individuals, such as when an author bequeaths a copyright to two children.

As is the case with other forms of property, joint ownership of copyright is legally termed a tenancy in common. Each joint owner is presumed to own an undivided proportional interest in the entire work; for example, if there are three joint owners, each is presumed to own one-third of the entire work.

Works Made for Hire. In many instances one person has created a work at the behest of another. In such circumstances, the creator does not own the copyright. As an easy example, consider a company that manufactures an appliance and has one of its employees write an instruction manual for the appliance. Logically, the company, and not the employee, should own the copyright.

Such situations are governed by the work-made-for-hire doctrine of the Copyright Law. Under the Copyright Law, copyright ownership vests initially in the author of the work. In cases of works made for hire, the law specifies that the employer or other person for whom the work is prepared is deemed to be the author (9). Thus the appliance company would be deemed to be the author, and hence the initial copyright owner, of the copyrighted instruction manual.

The law specifically defines what is and is not a work made for hire. It allows for two, and only two, possibilities.

First, a work made for hire is a work prepared by an employee within the scope of his or her employment. As the law does not define employee or employment, there were differences of opinion over the meaning of these terms until the Supreme Court resolved the matter. Employment, the Court ruled, has the same meaning as that commonly understood under the law of agency (10). Thus, although many factors determine employment, eg, the method of payment for services, withholding of taxes, location where work is performed, supplier of tools and instruments, duration of engagement, etc, it is safe to say that a relatively formal and commonly understood employment relationship is required, as distinguished from a situation of special commission or of independent contracting. However, in a work-made-for-hire situation, the parties may nevertheless agree in writing that the work is not a work made for hire.

Second, a work may be a work made for hire if it is specially ordered or commissioned and meets both of the following requirements: (1) it must be a contribution to a collective work, part of a motion picture or other audiovisual work, translation, a supplementary work, a compilation, an instructional text, a test, answer material for a test, or an atlas; and (2) the parties must expressly agree in a written instrument signed by them that the work is to be considered a work made for hire.

Because the circumstances under which an independently commissioned work is considered a work made for hire are limited, the commissioning party is likely to seek a transfer of copyright in the form of an assignment or license.

Transfers and Licenses of Copyright. Like other forms of property, copyright may be freely transferred. However, there are certain special rules governing the transfer of copyrights, and certain aspects of the law concerning transfer of property are of special importance to copyright.

Ownership of the copyright in a work is distinct from ownership of the material object, ie, the copy or phonorecord, in which the copyrighted work is embodied. The transfer of one does not constitute transfer of the other. For example, if a painter sells his or her painting, ie, the material object, such as canvas and oils, the painter does not automatically transfer the copyright in it; sale of that copyright, so as to allow reproduction of the oil painting in printed posters, does not transfer the material object.

It has often been said that copyright is a bundle of many different rights. There are many different ways in which a particular copyright may be exploited. The law allows copyright ownership to be virtually infinitely divisible. That is, each of those rights, in any subdivision conceivable, may be sold separately. The owner of any particular exclusive right is deemed to be the owner of copyright for that right. Thus, for example, the author of an article may sell the exclusive right of first publication of that article, but nothing else, to another, and that will result in two owners of copyright in that article, ie, the purchaser of the right of first publication, who will own only that right, and the author, who will own all other rights.

As a practical matter, an important distinction must be made between two methods by which copyrights are exploited. In the case of assignments or transfers of ownership of the copyright, either in whole or in part, the purchaser becomes the outright owner of the copyright or the particular right at issue. Alternatively, license of copyright, either in whole or, more likely, in part, is merely the permission to use the copyrighted work in the particular manner specified. Although exclusive licenses constitute transfers of copyright ownership for the particular rights involved, nonexclusive licenses do not. The distinction is important because the Copyright Act requires that transfers of copyright ownership must be in writing to be valid, whereas nonexclusive

licenses need not be in writing.

In the case of joint ownership of works, where the joint owners are treated as tenants in common, each co-owner may only transfer his or her own interest in the copyright and not the co-owner's interest. Thus a co-owner may not grant an exclusive license, which constitutes a transfer of copyright ownership, without the co-owner's permission. However, any co-owner may grant a nonexclusive license to use the copyright without the co-owner's permission. In this case, the co-owner granting the license must account to all other co-owners for their proportional shares of any profits realized by the nonexclusive license.

The Copyright Law provides a termination right to authors or, if they are dead, to their surviving spouses and children (11). Any transfer of copyright made after January 1, 1978, by an author may be terminated between 35 and 40 years after the transfer is made, and the copyright recaptured. The technical formalities concerning such terminations are intricate.

Copyright Formalities

Changes to the Copyright Law, starting with the 1976 Copyright Act and continuing with the Berne Convention Implementation Act of 1988, have radically changed U.S. copyright law regarding copyright formalities; many formalities previously of paramount importance have been eased or entirely eliminated.

Copyright Notice. In the past, the law contained an absolute requirement that each copy of a published work bear a proper copyright notice. Failure to comply with the technicalities of the law's notice provisions resulted in the unintentional loss of protection for many works.

An amendment to the law abolished the notice requirement for all works published on or after March 1, 1989. For such works, no copyright notice is required. Notice, however, is still required on all copies of works first published before that date. In addition, notice is still widely used even when it is not required so as to inform the world of the copyright status of the work. Notice consists of the symbol © (the letter *c* in a circle), the word *Copyright*, or the abbreviation *Copr.*; the name of the copyright owner; and the year of first publication.

Although copyright notice is no longer a prerequisite to copyright protection, it is still valuable as a nonlegal matter. It serves to warn off infringers and to identify the copyright owner to those seeking a license.

Copyright Deposit. The law requires that copies of every published work be submitted to the United States Copyright Office, which is a branch of the Library of Congress. The purpose of this requirement is to stock the shelves of the library. This requirement is usually satisfied as part of the registration process; failure to make deposit may ultimately lead to a fine, but will not affect the existence of the copyright.

Copyright Registration. The term *copyrighting a work* is usually misused. Although it refers to registering a work with the Copyright Office, copyright registration is not required for copyright protection. Federal copyright protection exists from the moment a work is created, ie, fixed in a tangible medium of expression, even if it is never registered.

Although copyright protection is not dependent on registration, registration does have important procedural advantages. First, in the case of works of U.S. authors, no lawsuit for copyright infringement may be brought until a work is registered. Second, many important remedies in a lawsuit, such as recovery of statutory damages and attorneys' fees, are not available to a copyright owner unless registration has preceded the infringement; there is a three-month grace period from the publication date for published works. Third, the certificate of copyright registration provided by the Copyright Office is *prima facie* evidence of the facts it contains and shifts the burden of proof concerning those facts from the copyright owner to the defendant in a lawsuit.

Copyright registration is easily accomplished. The copyright claimant completes a relatively simple form, returns it to the Copyright Office with a nominal fee, and deposits copies of the work. Special provisions allow for nondisclosure of trade secrets or full computer programs and the like in the deposit.

Transfers of copyright ownership must be in writing to be valid. The Copyright Office records any documents pertaining to copyrights, including transfers. Such record-keeping sometimes has important consequences, as in the perfection of security interests in copyrights.

Copyright Duration. Two different regimes of copyright duration apply in the United States. One is for works first created, published, or registered for copyright on or after January 1, 1978 (new law works); the other is for works published or registered before that date (old law works) (12). In all cases, copyright terms run through December 31 of their anniversary year.

New Law Works. The basic copyright term for new law works is the life of the author plus 50 years after the author's death. In the case of joint authors, the life in question is that of the longest surviving author.

If the duration of the author's life is not known, that is, for anonymous and pseudonymous works and works made for hire, the term is 75 years from publication or 100 years from creation, whichever expires first.

Old Law Works. Protection for works registered or published prior to 1978 endures for a dual-term system. There is an initial term of 28 years from the earlier of publication or registration, followed by a renewal term of an additional 47 years, for a total of 75 years of protection. Under a 1992 amendment to the Copyright Act, renewal is automatic, but there are procedural advantages to obtaining a renewal registration in the last year of the initial term.

The law contains a complicated provision, which case law has further elaborated, concerning ownership of renewal rights. Generally, renewal rights do not vest until the last year of the initial term. At that time they vest in the author. If the author is dead, the rights vest in the author's surviving spouse and children as a class; in the author's executor, ie, the beneficiaries under the author's will, if there is no surviving spouse or children; or in the author's next of kin under applicable state law if the author did not leave a will.

As a general rule, these renewal rights may be assigned in advance, but such advance assignments are only binding if those making them survive to the time when renewal rights vest (13,14). For example, if an author assigned renewal term rights in advance of renewal, but died during the initial term of copyright leaving a surviving spouse at the time of renewal, the surviving spouse would own the renewal rights and the author's assignment of the renewal term rights would be ineffective.

Copyright Rights

The Copyright Act grants copyright owners five exclusive rights (15). These rights include not only the right to do specified actions, but also to authorize them.

The Right to Reproduce in Copies. The most basic copyright right, of course, is the right to reproduce the copyrighted work in copies or phonorecords. This is the right to prevent unauthorized duplication of the work, such as the printing of an article without the copyright owner's consent.

The Right to Prepare Derivative Works. Many copyrighted works serve as the basis for derivative works, in which the underlying work is recast, transformed, or adapted. Examples include translations, motion pictures made from novels, and musical arrangements. Derivative works can be a significant source of income for copyright owners.

The Right of Public Distribution. The exclusive right to reproduce the copyrighted work also entails public distribution of copies, by sale or other transfer of ownership. This right, too, is the copyright owner's.

The Right of Public Performance. Certain types of works, for example, musical compositions, plays, and choreographic works, are meant to be performed. Public performance of these works is the copyright owner's exclusive right.

The Right of Public Display. Other types of works, notably pictorial, graphic, and sculptural works, are meant to be displayed. Again, their public display is the copyright owner's exclusive right.

Author's Rights. The Copyright Law was amended on June 1, 1991, to grant very limited additional rights to authors of certain types of works, even if they had parted with copyright ownership (16). These author's rights, sometimes referred to as moral rights, are applicable only to works of visual art that exist in single copies or multiples of up to 200. Even within this limited category of works, there are many exceptions, for example, author's rights do not apply to works made for hire. The author's rights are those of attribution, that is, right to have the author's name attached to or deleted from the work, and integrity, that is, the right to prevent mutilation or distortion of the work that would prejudice the author's honor or reputation.

Limitations and Exemptions

Fair Use. The Copyright Law contains several limitations on copyright rights and exemptions for certain uses. The best-known exemption is the fair use doctrine. Certain uses of copyrighted works that would otherwise be infringements are excused from liability because they are fair. The law gives examples of uses for purposes such as criticism, comment, news reporting, teaching, scholarship, and research. Each case must be judged on its particular merits and facts, however, and no particular use is presumed to be a fair use.

The Copyright Act requires the courts to consider at least four factors in each case: (1) The purpose and character of the use, including whether it is of a commercial or nonprofit educational nature. Commercial use is presumptively not fair use, although that presumption may be overcome, for example, for legitimate parodies that do not take too much of the copyrighted work. (2) The nature of the copyrighted work; it has been held, for example, that use of an unpublished work is less likely to be a fair use than use of a published work, and use of a scholarly or scientific work is more likely to be a fair use than use of works of pure entertainment. (3) The amount and substantiality of the portion used in relation to the copyrighted work as a whole; the less used and the less significant the portion used, the more likely that fair use will be found. (4) The effect of the use on the potential market for or value of the copyrighted work. This factor has been said to be the most significant fair use factor, because any real harm to the copyrighted work or its exploitation defeats the purpose of copyright protection (17).

The fair use doctrine has given rise to voluminous case law and commentary. Recent Supreme Court decisions have included *Harper & Row Publishers, Inc. v. Nation Enterprises*, 471 U.S. 539 (1985); *Sony Corp. v. Universal City Studios, Inc.*, 464 U.S. 417 (1984); and *American Geophysical Union v. Texaco, Inc.*, 802 F. Supp. 1 (S.D.N.Y. 1992).

First Sale Doctrine. Although the copyright owner has the exclusive right to distribute copies to the public, the bona fide possessor of a particular copy, in most circumstances, may further dispose of that copy without the copyright owner's consent. Thus, for example, the purchaser of a book may freely resell that copy. Hence, used book stores do not violate copyright law. Similarly, a copy of a work legitimately owned may be displayed publicly, as in the case of a picture hung in a museum. The Copyright Act has been amended to prohibit the rental of sound recordings or computer software, even though that rental would have been permitted by the first sale doctrine. Frequently, computer software is sold under a so-called license, which purports to bind the purchaser to limitations on the use and transfer of the software beyond the requirements of the law.

Infringement

Anyone who violates the exclusive rights of a copyright owner is liable for infringement in a lawsuit brought in federal court. There is a three-year statute of limitations on copyright infringement actions.

The Test for Infringement. It is rare that actual evidence of copying exists. Proof of copying is usually circumstantial and is shown by a two-part test. First, the alleged infringer must be shown to have had access to the copyrighted work. Second, the two works must appear to their hypothetical intended audience to be substantially similar. The two parts of the test may be seen to be in balance; while both must be present, the greater the evidence of substantial similarity, the less evidence of access is necessary, and vice versa.

Infringement of rights other than that of reproduction, such as the rights of public performance or display, are more easily proven directly by evidence of the infringing acts, eg, by a tape recording of a public performance.

Remedies. A copyright owner successfully proving infringement has three remedies available: recovery of monetary damages; injunctive relief; and recovery of costs, including attorneys' fees.

Damages may be recovered in one of two ways, at the choice of the copyright owner. First, the copyright owner is entitled to actual damages and to the infringer's profits resulting from the infringement. Because these measures of damage are often difficult to prove, the law alternatively allows for statutory damages, provided that copyright registration has been timely made. Statutory damages of between \$500 and \$20,000 for each work infringed, and not for each act of infringement, are assessed at the discretion of the court. The limits may be lowered to \$200 for truly innocent infringement, or raised to \$100,000 for willful infringement.

Injunctive relief, making the infringer stop infringing, is often more important to the copyright owner than recovering damages. The court may craft appropriate injunctive relief.

Within the court's discretion, and subject to timely registration, the prevailing party in an infringement suit may be awarded the costs of the litigation, including a reasonable attorney's fee.

International Copyright

Because copyright easily transcends national boundaries, several copyright conventions have been developed to protect copyrights internationally. The best known and most widely effective conventions are the Berne Convention for the Protection of Literary and Artistic Works and the Universal Copyright Convention; the United States is a signatory to both. In varying degrees, the treaties specify minimum standards that each member country's copyright law must meet.

International copyright treaties follow the principle of national treatment. Each member country treats nationals of other countries at least as well as it treats its own nationals.

BIBLIOGRAPHY

"Trademarks and Copyrights," in *ECT* 3rd ed., Vol. 23, pp. 348–374, by J. A. Baumgarten and C. H. Lieb, Paskus, Gordon & Hyman.

1. U.S., *Constitution*, Art. I, sec. 8, cl. 8.
2. 17 U.S.C. §§101 *et seq.* (1988).
3. 17 U.S.C. §102(a) (1988).
4. *Feist Publications, Inc. v. Rural Telephone Service Co.*, 113 L. Ed. 2d 358 (1991).
5. *Bleistein v. Donaldson Lithographing Co.*, 188 U.S. 239 (1903).
6. 17 U.S.C. §105 (1988).
7. 17 U.S.C. §102(b) (1988).
8. *Mazer v. Stein*, 347 U.S. 201 (1954).
9. 17 U.S.C. §201(b) (1988).
10. *Community for Creative Non-Violence v. Reid*, 109 S.Ct. 2166 (1989).
11. 17 U.S.C. §203 (1988).
12. 17 U.S.C. §§302–305 (1991).
13. *Fred Fisher Music Co. v. M. Witmark & Sons*, 318 U.S. 643 (1984).
14. *Stewart v. Abend*, 110 S. Ct. 1750 (1990).
15. 17 U.S.C. §106 (1988).
16. 17 U.S.C. §106A (1991).
17. U.S.C. §107 (1988).

General References

N. Boorstyn, *Copyright Law*, Lawyers Co-Operative Publishing Co., New York, 1981.

P. Goldstein, *Copyright*, Little Brown & Co., Boston, 1989. A new treatise by a highly regarded copyright scholar.

B. Kaplan, *An Unhurried View of Copyright*, Columbia University Press, New York, 1967.

M. B. Nimmer and D. Nimmer, *Nimmer on Copyright*, Matthew Bender, New York, 1991, 4-vol treatise. The most often cited and detailed copyright treatise, updated semi-annually.

W. Patry, *Latman's The Copyright Law*, 6th ed., Bureau of National Affairs, Inc., Washington, D.C., 1986. An excellent single-volume treatment.

Government Publications

The United States Copyright Office furnishes a wide variety of circulars and publications on all aspects of copyright, as well as registration forms. All may be obtained from the United States Copyright Office, Library of Congress, Washington, D.C., 20559.

Periodicals

A great many legal periodicals and newsletters contain scholarly articles on copyright or reports of recent developments. Among the best are the following.

ASCAP Copyright Law Symposium, Columbia University Press, New York. An annual collection of award-winning articles on copyright.

N. Boorstyn, *Copyright Law Journal*, San Francisco. A newsletter that analyzes recent case law.

Copyright Law Reporter, Commerce ClearingHouse (CCH), Chicago. A monthly service reporting on recent legislative and regulatory developments and all cases relating to copyright.

Journal of the Copyright Society of the U.S.A., published quarterly by the Copyright Society of the U.S.A., Center for Law and the Arts, Columbia University School of Law, New York.

Patent, Trademark and Copyright Law Journal, Bureau of National Affairs, Washington, D.C. A weekly newsletter of current developments.

History of the 1909 and 1976 Copyright Acts

Because of the complexity of the law, the legislative histories of the 1909 and 1976 Copyright Acts are vital to an understanding of the intent of the statutes. These histories are sufficiently complex that they have generated secondary works to aid in their reference.

E. Brylawski and A. Goldman, eds., *The Legislative History of the 1909 Copyright Act*, Fred B. Rothman & Co., Hackensack, N.J., 1976. A complete reprinting of the legislative history of the 1909 act.

A. Latman and J. Lightstone, eds., *The Kaminstein Legislative History Project*, Fred B. Rothman & Co., Littleton, Colo., 1981. A concordance and index of the legislative history of the 1976 act.

Specialized Subjects

Copyright Laws and Treaties of the World, Unesco, Paris, and Bureau of National Affairs, Washington, D.C., 1990. The complete text of every country's copyright statute.

P. Geller, *International Copyright Law and Practice*, Matthew Bender, New York, 1990. A useful country-by-country review.

W. Patry, *The Fair Use Privilege in Copyright Law*, Bureau of National Affairs, Washington, D.C., 1985. A comprehensive and much-cited review of fair use.

I. Fred Koenigsberg
White & Case

TRADEMARKS

A trademark, as defined in the federal statutes, includes any word, name, symbol, device, or combination thereof, adopted and used by a manufacturer or merchant to identify his or her goods and distinguish them from those manufactured or sold by others. Related to trademarks are service marks, certification marks, and collective marks. Service marks are names or symbols used to identify the services of one party and distinguish them from the services of others. Certification marks are used in connection with the products or services of one or more parties other than the owner of the mark to certify the product's origin, material, mode of manufacture, quality, accuracy, or other characteristics, or that the work or labor on the product was performed by members of a union or other organization. Collective marks are trade or service marks used by members of a cooperative, association, or other collective group. They include marks indicating membership in a union, association, or other organization. Principles applicable to service, certification, and collective marks are generally similar to those applicable to trademarks.

Ownership of a trademark confers the exclusive right to use it or authorize its use in connection with the goods of its owner or goods made by others to the owner's quality standards. Conversely, it also confers the right to prevent others from using the mark, or another mark so similar as to be confused for the original, on similar or related goods. Trademark protection is a type of restraint afforded by the courts against unfair competition.

In the United States and other common law countries, the right to protection of a trademark arises independently of statutory provisions. Both the public and the owner have an interest in such protection. The owner's interest lies in protecting the value of the mark as a device for attracting and retaining customers for the goods sold under the mark, based on recognition gained through use of the mark in advertising and the reputation for quality of the goods on which it is used. The public interest lies in avoiding the confusion, deception, and fraud that would be likely as a result of unauthorized use of the mark by others.

Although distinct from trademarks, trade names also are entitled to protection against unfair competition. Trade names and trademarks should not be confused. A trademark identifies the particular goods of a business; a trade name identifies the business itself. Trade names include individual names and surnames, firm names, and other names used by manufacturers, industrialists, merchants, agriculturists, and others to identify their businesses, vocations, or occupations. Depending on the manner of its use, a trade name may also function as a trademark. For instance, the name General Electric, when used to refer to the corporation, is a trade name; but when applied to merchandise, eg, General Electric refrigerators, it is a trademark.

Trademarks, Patents, Copyrights. A trademark is acquired by use on or in connection with the goods of its owner. It is accorded legal protection at common law independent of statutory provisions or registration. Registration, available under the laws of the states as well as under federal statute, enhances or facilitates protection of the trademark right. Acquisition of the trademark right by use in connection with the goods is a prerequisite to obtaining registration. Under the 1988 Trademark Revision Act, applications for trademark registration may be filed on the basis of a bona fide intent to use the mark, but proof of actual use is still required before a registration is issued.

Patents afford the owner the right to exclude others from making, using, or selling an invention, and are entirely dependent on statutory registration. They are acquired by disclosing an invention in an application duly filed and prosecuted in accord with the patent laws (see PATENTS AND TRADE SECRETS).

Copyrights afford in most cases an exclusive right to control distribution, reproduction, adaptation, public performance, and public display of literary or artistic works. They arise automatically upon creation of an eligible work, but the exercise of such rights is governed exclusively by federal statute.

Authority for protection of patents and copyrights is set out in the Constitution and is the exclusive province of the federal government. Federal legislation to protect trademarks is based on the authority of Congress, under the commerce clause of the Constitution, to regulate interstate and foreign commerce of the United States; protection afforded by individual states is based on their power to regulate intrastate commerce.

History

Trademarks were known in ancient civilizations. Wine jars found in Pompeii bore the mark Vesuvinum, identifying their contents as a wine produced in the region of Vesuvius. The artisan's guilds of Europe adopted marks to identify merchandise originating with their members; such marks were protected against falsification by penal laws.

British and U.S. courts began developing case law to protect trademarks against infringement in the early 1800s. In 1883, the United Kingdom adopted a statute that provided for registration of fancy words as trademarks. The United States enacted its first federal trademark statute in 1870, but the statute was declared unconstitutional in 1879. Subsequent federal trademark statutes were adopted in the United States in 1881, 1905, and 1920; the present comprehensive statute, known as the Lanham Act, was enacted in 1946. The Lanham Act was substantially revised for the first time in 1988 by the Trademark Law Revision Act of 1988. The 1988 act, which became effective on November 16, 1989, both modified and supplemented the earlier statute.

Selection of Trademarks

Trademarks distinguish the merchandise of the owner of the mark from the merchandise of others. Any designation commonly used by the public in describing the merchandise, its functions, or its properties cannot be preempted by one manufacturer to the exclusion of the public. Thus a mark for which exclusive protection is sought cannot be merely descriptive of the merchandise. Preferably, a mark should be fanciful or arbitrary. A trademark must also be adequately distinguishable from other trademarks for similar merchandise to avoid confusing or deceiving customers about the origin of the goods (1).

To determine whether a prospective mark is distinct enough from other marks, a search in the fields of commerce in which the mark's use is contemplated is customary. Since most marks are registered under the federal statute in the United States Patent and Trademark Office, this register provides a principal field of search. The search can be made in the Patent and Trademark Office under a classification of the goods for which marks are registered. A complete search, however, also requires a search of marks registered under state laws and of marks not registered but in use as indicated in various trade directories, telephone books, and the like. A number of professional organizations provide thorough searches including these additional fields, and use of their services is highly advisable prior to adoption of a new mark. In some instances, an international search may also be advisable; professional search organizations are capable of performing these international searches as well.

The Trademark Register of the United States, an annual publication by the Trademark Register, Washington, D.C., is a useful means of determining whether or not a mark under consideration for adoption, or any mark resembling it, has already been adopted for goods of a similar character. The annual editions list the marks on the *Federal Register*, ie, marks registered in the United States Patent and Trademark Office, in alphabetical order under the various classes of merchandise and services together with the registration number and date. Full details concerning the goods, application serial number, date of application, and alleged date of first use with respect to a particular registration can be obtained by ordering a copy of the registration from the United States Patent and Trademark Office. The Patent and Trademark Office also issues a weekly publication, the *Official Gazette*, which lists all registrations numerically for the week in which they are issued. The *Official Gazette* is available in many libraries. Other reference works that may be helpful in a preliminary survey are the *Compu-Mark Directory*, *Merk Index*, *Thomas Register*, *Physicians' Desk Reference*, and the *Condensed Chemical Dictionary*.

Principal considerations used to determine whether a prospective mark conflicts with an existing third-party mark include the similarity of the marks and the relationship between the kinds of merchandise for which the marks are used or are intended to be used (2). Without some kind of relationship

between the goods to which the marks are applied or are intended to be applied, even identical marks would be unlikely to cause confusion. Similarity between marks may exist in appearance, phonetics, connotation, or any combination of these qualities. Similarity of an accented or initial syllable, or otherwise dominant portion of a mark, may be an important factor in such determinations. However, in the final analysis, confusing similarity is determined with respect to the marks as a whole. For example, in a 1989 decision, the federal circuit court affirmed the Patent and Trademark Office Trademark Trial and Appeal Board's determination on summary judgment that the mark Pecan Sandies is not confusingly similar to the mark Pecan Shorties, in each case for pecan cookies. The court stated that the term *pecan* was merely descriptive of the principal ingredient of both parties' cookies and that, taken as entirety, they did not convey the same impression (3).

The *Trademark Reporter*, a publication of the United States Trademark Association, carries in its monthly issues and yearly index a revealing tabulation of marks involved in decisions of the Patent and Trademark Office and the courts, and found to be confusingly similar as applied to the goods for which they were intended.

The nature of the goods for which the mark is used or intended to be used is an essential element in determining the likelihood of confusion between otherwise similar marks. If goods bearing similar marks are of such a nature that a purchaser or prospective customer would be likely to assume that the goods have a common source of origin, then a basis for confusion exists.

For example, the mark Classic Tiffany for luxury automobiles was held confusingly similar to the mark Tiffany for luxury items, including jewelry and household goods; Gentle Touch for deodorant was held confusingly similar to Kind Touch for a cleansing preparation; Calvins for condoms was held confusingly similar to Calvin and Calvin Klein for men's apparel and men's toiletries; Avita was held confusingly similar to Aveda, both for hair care products; and Brador for malt liquor was held confusingly similar to Bras d'Or for cognac. On the other hand, Lexus for a new line of cars was not considered confusingly similar to Lexis for computer-assisted legal research; product difference was decisive in that case.

Other types of goods held to be such as to give rise to the likelihood of confusion included soft drinks and containers; beauty preparations and soaps, cleansers, and lotions; doors and chairs; lighting fixtures and furniture; and guns and steel products. The degree of fame enjoyed by a mark can add to its scope, bringing a broader range of goods and services under its protection.

In addition to the question of possible conflict with other trademarks, it is also necessary to consider whether a proposed new mark has such descriptive character with respect to the goods for which it is intended as to render the mark incapable of distinguishing the goods of the owner from such goods in general. At one extreme are generic terms, ie, names that the public uses to identify the product itself, rather than the source of the product. Such terms can never function as trademarks, at least not for the products they denote, since they are in the public domain, free for all to use. At the other end of the spectrum are terms that are descriptive, but are used long enough and extensively enough such that the public identifies them as trademarks. The latter, although descriptive, may nevertheless function and be protected as trademarks on the ground that they have acquired a secondary meaning in the mind of the public.

Descriptive marks that have been held registrable include Lens Bright for eyeglass-cleaning fluid; Beer Nuts for sweetened, salted nuts; Home Savings for savings and loans services; Mastic for vinyl siding; The Daily Catch for a seafood restaurant; and California Cooler for a wine cooler. Marks held incapable of trademark function include Intelligent Modem for a computer modem; Honey Roast for nuts roasted with honey; Shear Pleasure for a hairdressing salon; Cozy Warm Energy Savers for flannel pajamas and nightgowns; and Pizzaz for zippy or zesty pizza.

Names of persons, especially surnames, are treated in the same manner as descriptive terms, they may function as trademarks only if they have been used long enough and extensively enough such that the public identifies them as trademarks. When used exclusively for merchandise of a given class to such an extent that the public understands them as trademarks, they acquire secondary meaning and are accorded protection. For instance, Ford and DuPont have long since been recognized as trademarks, despite the fact that they are surnames. Other names or parts of names that have become trademarks are Gallo, Gucci, Calvin Klein, Sardi's, and McDonald's.

Marks that are not descriptive, but rather suggest the character, quality, or function of the goods are much favored as they are especially useful in advertising and encourage easy recall on the part of the public. Examples of suggestive marks include Ecotrin for an enteric-coated aspirin; Citibank for a modern or urban bank; Classic Cola for a soft drink; and Sheer Elegance for pantyhose. Such marks are entitled to protection in the first instance without acquisition of secondary meaning.

Depending on the circumstances, seemingly descriptive marks when taken as a whole might constitute only suggestive terms. A 1968 decision of the Court of Customs and Patent Appeals is illustrative. Sugar and Spice was held registrable for bakery goods. The court stated:

A mark that merely denotes ingredients, quality, or composition of an article is not capable of being adopted exclusively and used as a trademark, since for policy reasons, descriptive words must be left free for public use. On the other hand, a legal distinction—albeit often obscure—has been drawn between terms that are [merely descriptive and those that are] only suggestive of the goods. The terms sugar and spice, used individually, are well known and well understood by the purchasing public. However, when combined and used on bakery goods, we think they may function as an indication of more than a mere description of the goods. (4)

The final class of marks comprises those that are wholly arbitrary with reference to the goods for which they are to be used. They may be words having a meaning that bears no relationship to the goods for which they are used, eg, Apple for computers, Shell for gasoline, Arrow for liqueurs, and Mustang for a motel; they may be synthetic terms having no significance, eg, Kodak for cameras, Yuban for coffee, or Zazu for hair care services and products. These marks have the advantage of being least likely to lose their trademark status by becoming the generic name of the product and thereby entering the public domain.

In addition to letter combinations or numbers, such as 4711 on cologne and IBM for business machines, a trademark may be an acronym; a nickname such as Pan Am; a logotype, design, or symbol, such as the McDonald's golden arches or the Olympics' five interlocking rings; the configuration or contour of a product or its container, such as Lifesaver candy, the Perrier bottle, and Ferrari automobiles; a series of notes such as the NBC signature; or a song such as General Motors' advertising theme, "The Heartbeat of America."

Trademark Registration

Trademarks function may be characterized as an identification of origin, a guarantee of quality, and an advertising device. In identifying origin, a trademark affords a convenience to both owner and customers by minimizing confusion of the goods of the trademark owner with those of others. In guaranteeing quality, it assures the purchaser that the goods bearing the mark will have the same quality as those previously purchased under the same mark. In both instances, the public interest is cited by the courts as a basis for protection to avoid deception or confusion. In advertising, the trademark offers a convenient shorthand device to help focus the attention and the good will of the public on the product. Protection of the trademark represents protection by the courts of a property right of the trademark owner.

Registration Rights. The right to trademark protection arises through exclusive use of the mark in connection with the goods of the owner. This right is protected by common law independent of registration. Trademark registration is a statutory creation; it affords a means of publicizing a claim to a trademark right and facilitating its protection. The federal statute, the Lanham Act, provides for two separate systems of registration: a Principal Register for marks fulfilling all requirements for full registration, and a Supplemental Register for those lacking certain of the requirements for registration on the Principal Register. Registration on the Principal Register constitutes constructive notice of the claim of ownership of the mark. Federal registration on either register extends protection to all parts of the United States and its territories and possessions. A common law trademark right, on the other hand, is effective only in the region where the mark has become known through use. Federal registration on either register also confers jurisdiction of trademark actions on the federal courts and affords the trademark owner additional rights against infringers. Substantial advantages can thus be derived from federal

trademark registration.

A trademark is registrable on the Principal Register if it is in use in commerce subject to regulation by Congress, ie, interstate and foreign commerce of the United States, and does not (1) comprise immoral, deceptive, or scandalous matter, or matter disparaging persons, institutions, beliefs, or national symbols; (2) comprise the flag or coat of arms or other insignia of the United States or any state, municipality, or foreign nation; (3) comprise the name, signature, or portrait of an individual without his or her written consent, or the name, signature, or portrait of a deceased President of the United States during the lifetime of his widow without her written consent; (4) comprise a mark that so resembles an already registered mark, or one previously used in the United States and not abandoned, as to cause confusion or mistake or to deceive when applied to the goods of the applicant unless limitations of use can be implemented to prevent confusion, in which case concurrent registration may be permitted; (5) consist of a mark that, when applied to the goods of the applicant, is merely descriptive or deceptively misdescriptive, or is primarily geographically descriptive or deceptively misdescriptive, or is primarily a surname (5). Exceptions to the last item are permitted if the mark has become distinctive of the applicant's goods in commerce by use. Proof of substantially exclusive and continuous use of a mark in commerce over a period of five years may constitute *prima facie* evidence that the mark has become distinctive as used in connection with the applicant's goods.

The statute expressly provides that the trademark may consist of any symbol, label, package, configuration of goods, name, word, slogan, phrase, surname, geographical name, numeral, device, or any combination of any of the foregoing, provided it is capable of distinguishing the applicant's goods or services. A mark lacking the distinctiveness to qualify for registration on the Principal Register, but accorded registration on the Supplemental Register, may be registered on the Principal Register at a future date if it can be shown that the mark has become distinctive of the applicant's goods. Registration on the Supplemental Register, as distinguished from the Principal Register, does not constitute constructive notice of the claim of ownership by the registrant. Like registration on the Principal Register, registration on the Supplemental Register may serve as a basis for obtaining registration in countries that require home registration as a prerequisite. In addition, it provides a priority date for purposes of soliciting registration in foreign countries under international treaties (6).

Registration Procedure. The procedure for registering a trademark under the Lanham Act involves filing an application with the Patent and Trademark Office requesting registration on either the Principal Register or the Supplemental Register, and declaring either an actual, present use of the mark in interstate or foreign commerce or a bona fide intention to use the mark in commerce. Prior to the amendments of the 1988 act, a domestic applicant was required to assert actual, demonstrable use of the mark in commerce in connection with the goods or services for which registration of the mark was sought and could not apply for registration on the basis of intended future use. Although actual use still serves as a basis for filing an application for registration and is still a prerequisite for obtaining the registration itself, an application may now be filed prior to actual use so as to secure a mark and prevent a third party from acquiring rights to the mark in the interim.

An application filed on the traditional basis of actual use must set forth the kind of commerce, interstate or foreign, in which the mark is in use; the date of first use and the date of first use in commerce subject to regulation by Congress, ie, interstate or foreign commerce; the goods on which the mark is used; and the manner in which it is applied to the goods. The application must include a drawing that shows the mark in the form in which it is used and five specimens, ie, labels or other reasonable facsimiles such as photographs, that show the mark in the form in which it is applied to the goods. When specimens such as labels are not available or the manner in which the mark is used does not lend itself to inclusion of specimens in the application, photographs or other facsimiles may be used. In the case of a word mark, or letters or numerals, for which the design or form of the letters or numerals is not critical, the drawing may be typed, showing the mark in block letters. The application must be executed by the owner. The application is filed in the Patent and Trademark Office and must be accompanied by a filing fee (7). If an attorney is to prosecute the application, that is, respond to any objections of the Patent and Trademark Office, file and argue appeals, defend against opposition and cancellation proceedings instituted by others, and so on, a power of attorney should also be included.

After examination, any objections, including rejection in view of prior registrations, are communicated to the applicant in an office action, requiring a response within a limited period. The response may involve amendment of the application, an argument to overcome the grounds for rejection, or a showing of further facts. Final rejections of applications may be appealed to the Trademark Trial and Appeal Board, if necessary or desired, as described in the following case of adversary proceedings.

When the application is found to be allowable, it is published in the *Official Gazette* of the Patent and Trademark Office, and within 30 days, or within such extension thereof as may be granted to a potential opposer, any person may file an opposition alleging the belief that he or she will be damaged by the registration, and the grounds of that belief. The application is then transferred to the jurisdiction of the Trademark Trial and Appeal Board, which fixes terms for taking testimony by the opposer and the applicant. The testimony is taken before an official, ie, a court reporter, authorized to do so, and representatives of the adverse party may cross-examine. The depositions and evidence are submitted by the reporter to the Patent and Trademark Office. A final hearing is held before the Trademark Trial and Appeal Board, which renders a decision based on the testimony presented by the parties. This decision may be appealed to the Court of Appeals for the federal Circuit, or the applicant may have a remedy by civil action in the federal district courts. If the parties are able to negotiate a settlement while the matter is pending, the opposition can be withdrawn.

If no opposition is filed, or if a decision favorable to the applicant is made in an opposition, the application is allowed, and the registration is issued.

Should a party fail to file an opposition against a published application within the time allowed, the objection can be filed after issuance of the registration. This involves a proceeding for cancellation on grounds similar to those available for opposition. Although the procedures are similar, the burden of proof, which falls on the party seeking cancellation, is more strictly construed, at least in theory, in a cancellation proceeding.

Pursuant to the amendments of the 1988 Act, an intent-to-use application may be filed in advance of actual use and, subject to approval, remain pending for a specified period of time until a statement of use is filed. Once the registration is issued, the applicant can refer to the date of filing of the intent-to-use application as the date of first use for purposes of establishing priority. This new provision was added in part to address the problems encountered under the prior law of the necessity of expending start-up costs before being able to ascertain with certainty whether an application for registration would be granted and of the possibility that another party could make actual first use of the same or a similar mark between the time a mark was searched and cleared and the time it was put into actual use.

To receive a filing date, an application made on the basis of intent-to-use must include "a claim of bona fide intention to use the mark in commerce," a description of the goods upon or in connection with which the applicant has the intention to use the mark and the mode and manner in which it is to be used, and a statement that the applicant is entitled to use the mark and that, to the best of the applicant's knowledge, no other person or entity has the right to use the mark or a confusingly similar mark. A drawing of the mark must be submitted with the application.

As in the case of an actual use application, a filing fee must be transmitted with the intent-to-use application and the application must be executed by the owner of the mark. If the applicant makes use of the mark at any time during the examination process, he or she may amend the application and convert it to a traditional, actual use application. If the application has not been amended to allege use by the time its examination has been completed and it receives a favorable examination, it is published in the *Official Gazette* and, unless a successful opposition proceeding is brought, the applicant receives a notice of allowance. From the date of the notice of allowance, the applicant has six months to file a verified statement of use with specimens demonstrating actual use of the mark in commerce. An automatic six-month extension is available, and upon a showing of good cause, four additional six-month extensions may be obtained. Thus it is possible for an applicant to secure a total of three years after receipt of the notice of allowance within which to obtain registration of a trademark.

All marks registered under the Lanham Act are subject to cancellation by petition of a third party at any time within the five years following registration. The same time period applies after publication of a notice in the *Official Gazette* converting a registration under a prior trademark statute to one entitled to protection under the Lanham Act. Such conversions are allowed except in those instances in which the mark has become the common descriptive name of the article or substance, has been abandoned, or is disqualified by one of the five criteria listed earlier precluding registration. Cancellation proceedings may be brought after expiration of the five-year period only in certain circumstances, such as if the mark has become a generic name, has been abandoned, or was fraudulently obtained.

Marks registered only under prior trademark acts are subject to cancellation proceedings at any time at the initiative of someone who claims damage or potential damage by the registration of the mark. Certification marks can be similarly canceled if the owner fails to exercise control over their use or misuses them (8).

For maintenance of a registration, the Lanham Act requires that an affidavit or declaration under penalty of perjury be filed within the sixth year of registration showing that the mark is still in use in commerce subject to regulation by Congress or that its nonuse is the result of circumstances excusing nonuse and not to an intention to abandon the mark. If the document includes a statement that a mark has been in substantially continuous use for five consecutive years following the date of registration, the right of the registrant to use the mark for the same goods or services becomes incontestable. The registration of an incontestable mark is immune from cancellation on most grounds, but may still be subject to attack if the mark infringes a valid trademark right based on continuous use antedating the publication thereof, becomes the generic name of the goods, is abandoned, or was improperly registered. Pursuant to the 1988 Act, this statement of continuous use can now be filed before the five-year period of substantially exclusive and continuous use has expired. Thus the registration may acquire incontestable status soon after the expiration of the five-year period, rather than closer to the sixth year, as was typical under the prior law. Prior to the 1988 Act, the application for incontestability could not be made until after the five-year period had elapsed.

The requirement of filing a declaration within the sixth year of registration and the availability of incontestability apply only to marks registered under the Lanham Act, and those marks registered under earlier acts that have been converted to registrations under the Lanham Act by appropriate application and publication.

Marks registered only under earlier acts are not incontestable and do not require the filing of an affidavit of use within six years after registration or renewal. All marks, however, require renewal at ten-year intervals after registration. Such renewal is effected by filing a complete application within six months before expiration or within three months after expiration, identifying the goods or services for which the mark is still in use or showing that nonuse is the result of circumstances excusing such nonuse and not indicating an intent to abandon the mark. The nonuse of marks for alcoholic beverages during the period of prohibition was held to be excused in this sense, and the marks were upheld when prohibition was repealed. Nonuse has also been excused during wartime exigencies and, more recently, during the course of litigation over the right to use the mark.

Effect of Registration. Under the federal statutes, the registrant of a mark can bring civil action in the federal courts against the unauthorized use of a reproduction, counterfeit, copy, or colorable imitation of the mark in connection with any goods or services for which such use is likely to cause confusion or mistake or to deceive. A certificate of registration on the Principal Register of the Lanham Act, or republished from the act of 1881 or 1905, is *prima facie* evidence of the validity of the registered mark and the registration, the distinctiveness of the mark, the registrant's ownership of the mark, and the exclusive right of the registrant to use the mark in commerce.

If the mark has become incontestable because of proper filing of a statement of five years of continuous use, the registration is conclusive evidence of the validity and ownership of the mark and registration, of the distinctiveness of the mark, and of the exclusive right of the registrant to use the mark on or in connection with goods or services specified in the declaration conferring incontestability. Exceptions are made in certain well-defined circumstances, for example, fraudulent registration, abandonment, and generic use of the mark.

The registrant of a trademark can enforce his or her rights by civil action in federal court against one who makes an unauthorized use of the mark that is likely to cause confusion or mistake or to deceive. In cases of contributory infringement, where the unauthorized use is unknowingly made by one engaged solely in the business of printing the infringing mark or of reproducing for others an advertisement displaying the mark in a periodical or an electronic communication, the remedy is limited to an injunction against future printing or advertising of the infringing matter. Thus, although printers and publishers who innocently reproduce an infringing work may be liable for producing the infringing advertisements, labels, and so on, they do not suffer any monetary liability. The mark's owner can obtain only injunctive relief, that is, a prohibition against further printing or publication. Profits or damages are recoverable only if the infringement was committed with knowledge or an intent to cause confusion or mistake or to deceive.

A registrant prevailing in a federal court action against an infringer is granted an injunction that is enforceable in any District Court of the United States. In addition, the plaintiff may recover the defendant's profits, damages sustained by the plaintiff, and costs. The court may, in its discretion, assess up to triple damages. The court may also order the destruction of all labels, signs, prints, packaging, wrappers, receptacles, and advertising in the possession of the defendant bearing the infringing mark and all plates, molds, matrices, or the like for producing them.

Registration of a mark on the Principal (but not the Supplemental) Register, as well as registration under the Acts of 1881 or 1905, constitutes constructive notice of the owner's claim to the mark. Damages and profits are available only from the time of actual notice to an infringer, unless the owner displays with the mark as used a notice of registration in the form specified in the statute, that is, the phrase *Registered in United States Patent and Trademark Office*, the abbreviation *Reg. U.S. Pat. & Tm. Off.*, or the letter R enclosed in a circle. In those cases not precluded by the statute of limitations, damages and profits may be recovered for infringing acts that occurred within three years prior to institution of a civil action for infringement.

Registration of a mark on the Principal Register prohibits importation of merchandise that copies the registered mark. Under the regulations of the Department of the Treasury, the registration may be recorded with the Customs authorities and merchandise bearing the infringing mark is subject to seizure and exclusion. The recording procedure includes supplying the Bureau of Customs with a specified number of copies of the registration. These are distributed to the various ports of entry and border stations for reference and notice to the Customs inspectors (9).

Unlike trademarks, trade names are not registrable. However, the use of a trade name in a business in which such use is likely to cause confusion with a preexisting trademark is subject to remedies similar to those available against a confusingly similar trademark. Conversely, a preexisting trade name may serve as the basis for attack, by opposition, cancellation, or suit for unfair competition, on a confusingly similar trademark.

Under the International Convention for the Protection of Industrial Property, to which the United States is a party, an applicant for registration of a trademark in the United States is entitled to a right of priority in other member countries from the date of the U.S. application if a corresponding application is filed in the other country or countries within six months after the date of the U.S. application. Reciprocal rights are granted by the United States to nationals of other member countries. Nationals of such countries applying for U.S. registration based on a prior foreign registration need not allege use in commerce in the United States, but must allege a bona fide intent to use the mark in commerce in the United States. In such a case, the registrant may not recover for infringing acts prior to the date of the U.S. registration. Third party rights antedating the foreign priority date are not affected by the registration.

Transfer of Trademark Rights

A trademark and its registration may be assigned, but only together with the good will of the business in connection with which the mark is used or that part of the good will of the business connected with the use of and symbolized by the mark. An assignment of an intent-to-use application for registration is further limited to instances where the business in which the mark is to be used exists and is ongoing and is transferred to the assignee of the application along with the application and rights therein.

A trademark right does not exist apart from the good will of the business or part of the business in connection with which it is used. Accordingly, an attempt to assign the right without the associated good will is ineffective. In such a case, the owner divests himself or herself of the right and thus abandons the trademark. The intended assignee acquires no right based on prior ownership of the assignor. At best, an intended assignee may establish a new right to the mark based on his or her own exclusive use thereof, subject to prior rights of third parties.

An assignment of a trademark and its registration must be in writing and duly executed. A verification sworn to by the party signing the assignment before a notary or other official authorized to administer oaths creates a presumption that the assignment was validly executed. The United States Patent and Trademark office maintains a record of assignments submitted to it for recording. Unless recorded within three months of its date, an assignment of a registered mark is considered void if a subsequent purchase is made in good faith and the subsequent purchaser records in the Patent and Trademark Office first.

A trademark may be licensed, but special rules apply to such licensing because of the nature and function of the trademark right. Thus the public has an interest in the mark as a symbol of origin of the goods or services in the owner and should be able to rely on the mark as indicating the character and quality of the goods or services that the owner controls. Essential, therefore, to a valid license is provision for the licensor to control the nature and quality

of the goods or services on or in connection with which the mark is to be used. Not only does the licensor have the right to exercise such control, but control must be effectively exercised to validate the license. Without such an arrangement, the license is invalid and constitutes an abandonment of the trademark by the licensor, since the use by the licensee is inconsistent with the otherwise exclusive right of the owner to use the mark. On the other hand, when a license contains provisions for suitable quality control and such control is exercised, use of the trademark by the licensee benefits the licensor (10).

Proper Use of Trademarks

Proper use of a trademark by the owner or a licensee under a valid license is essential to preservation of the exclusive trademark right. The Lanham Act provides, for example, that nonuse for a period of two years constitutes *prima facie* abandonment of the trademark right. A trademark may also be lost through other circumstances. For instance, if the mark comes to be understood by the public, or the class of persons constituting the usual customers for merchandise sold under the mark, as a generic name for the product for which it is used, then the mark enters the public domain, and the exclusive right of the owner is lost. The danger of losing the exclusive right increases as a mark becomes more popular. Examples of marks that have been lost or severely endangered in this manner are Aspirin for acetylsalicylates, Escalator for moving stairways, and Cellophane for regenerated cellulose film. The Murphy bed, Yo-Yo toy, and Thermos insulated bottle were each originally trademarks that became generic names. Sometimes a mark that has become generic may be reestablished by an assiduous advertising campaign. Thus, for an earlier generation, Singer became largely synonymous with sewing machine, but through vigorous publicity it is now reestablished as a trademark.

In the drug field, there is a need for simple and useful nonproprietary names for drugs. The United States Adopted Names (USAN) Council chooses generic names for drugs used in the United States. A list is available (11).

Various usage rules have been devised to reduce the danger of a trademark degenerating into a generic name. Some companies publish information bulletins and management guides for the use of their trademarks (12). These rules may be particularly significant with respect to chemical products because their actual generic names are frequently cumbersome or not referred to in common parlance (13). It is essential that readily usable generic names be developed for any new chemical product and that the rules of proper use be followed in referring to the product and its trademark. The principle of these rules is to limit the use of the mark on goods or in advertising so as to signify the origin of the goods (with the trademark owner) not the goods themselves. Thus the mark is appropriately used as an adjective followed by a generic name for the goods. It should be set off from its context by capitalization, distinctive lettering, quotation marks, or the like. It should not be used as a noun, and most important, not as a plural noun. It should not be used as a verb or as a term characterizing a process. It should be clearly identified as a trademark in each context in which it is used. Thus, if it is registered under the federal statute, its trademark status may be indicated by a circled R or by the legend (or its abbreviation) showing registration in the United States Patent and Trademark Office. If the mark is not registered, its trademark status may be indicated by an adjacent *TM* or by the term *Trademark* related, if appropriate, by asterisk to the mark itself. Preferably, the name of the trademark owner also should appear. Emphasis on the trademark status of the mark wherever it is used in commerce is the best means of preventing degeneration of the mark into a generic name for the goods.

Questions may arise as to the propriety of use of a mark by parties other than the owner in connection with goods originating with the owner and originally sold under the mark. Dealers are free to use the mark in advertising the goods for sale to customers, but they are not free to use the mark as a designation for a business establishment, except to the extent that this may be permitted by license from the trademark owner. Use of the mark in the resale of used or reconditioned goods bearing the mark of the manufacturer is not objectionable if the used or reconditioned status is clearly indicated to purchasers. Use of an ingredient mark on a composition containing the ingredient but compounded by a party other than the owner of that mark has been held permissible, provided that no overemphasized display is given the ingredient mark and adequate legends are placed on the goods to show the proportion of the ingredient contained and to indicate that the composition has been manufactured by its producer and has no connection with the owner of the ingredient mark.

State Registration

Registration of a trademark under state law is appropriate when the mark is to be used primarily in intrastate commerce. Procedure for registration generally parallels, but is simpler than, federal procedure. The office of the Secretary of State is usually the agency charged with granting trademark registration. Rules relating to procedures for registering trademarks at the state level are established by state legislation and, as such, vary from state to state. State registration may be convenient in some cases, such as when an owner desires to license or franchise one party in each state of a given territory without extending rights under the mark beyond the state boundary.

A number of states afford remedies against encroachments that may exceed those afforded under the federal statutes. One such significant remedy is that against dilution of the mark, where use of a mark that is similar to the mark in question, although not confusingly similar because of a dissimilar or noncompetitive use, nevertheless tarnishes, degrades, or dilutes the distinctiveness of the mark (14).

Foreign Trademark Registration

The trademark laws of foreign countries differ in many fundamental respects from the laws of the United States. Countries deriving their system of laws from the United Kingdom generally recognize use of a mark as the basis for the trademark right. Registration is permitted in some of these countries based on an allegation of intent to use the mark, and continuance of the registration is dependent on demonstrating within a specified period of time that the intended use has occurred. Countries that derive their legal systems from continental Europe generally permit registration without use, although nonuse may be a ground for cancellation after a period of years.

Under the International Convention for Protection of Industrial Property, which has some 100 member countries, including the United States, if an application for registration of a trademark has been filed in one member country, a subsequent application for registration of the same trademark in another member country will be treated as if it were filed on the same date as the first application for purposes of priority, as long as the later application is filed within six months of the first and convention priority is claimed (15). A U.S. application may also be based on foreign registration or a foreign application for registration cojoined with a bona fide intention to use the mark in commerce in the United States.

Twenty-nine countries of the convention are also parties to a treaty for trademark registration known as the Madrid Arrangement. Under this treaty, a mark registered in the member home country of a trademark owner may be registered in the other member countries by filing with the International Trademark Bureau in Berne an application showing the home country registration. The bureau communicates the resulting registration to the registration authorities of the other member countries, which may accept or reject it in whole or in part. Subject to such limitations as are thereby imposed, the mark becomes effective as a registered mark in each member country under the provisions of local law governing registered trademarks. The term of a registration under the Madrid Arrangement is 20 years. The United States, Canada, the United Kingdom, and the Scandinavian countries are not members of the Madrid Arrangement, and registrants whose home registration is in these countries are not entitled to corresponding registration under the Madrid Arrangement. Members of the Madrid Arrangement include primarily continental European countries, their overseas possessions and former possessions, and several Asian countries.

BIBLIOGRAPHY

"Trademarks and Copyrights," in *ECT* 3rd ed., Vol. 23, pp. 348–374, by J. A. Baumgarten and C. H. Lieb, Paskus, Gordon & Hyman.

1. United States Trademark Association, *Trademark Management—A Guide for Executives*, New York, 1981, pp. 1–54.
2. J. T. McCarthy, *Trademarks and Unfair Competition*, 2d ed., Lawyers Cooperative Publishing Co., Rochester, N.Y., 1984, §§ 23:1, 24:1, pp. 42–44, 161–162.

3.

Keebler Co. v. Murray Bakery Products, 866 F.2d 1386, 9 U.S.P.Q. 2d, 1736 (Fed. Cir. 1989).

4.

In re Colonial Stores, Inc., 394 F.2d 549, 157 U.S.P.Q. 382 (C.C.P.A. 1968).

5.

Sec. 2, *Trademark Act of 1946*, as amended, 15 U.S.C. § 1052.

6.

Paris Convention for the Protection of Industrial Property of 1883, rev. at Washington, D.C. (1911), The Hague (1925), London (1934), Lisbon (1958), and Stockholm (1967).

7.

37 C.F.R. §§ 2.6, 2.21.

8.

Sec. 14, *Trademark Act of 1946*, as amended, 15 U.S.C. § 1064.

9.

Sec. 42, *Trademark Act of 1946*, as amended, 15 U.S.C. § 1124.

10.

J. T. McCarthy, *supra* note 2, pp. 830–831.

11.

U.S. Pharmacopeial Convention, Inc., *USAN and the USP Dictionary of Drug Names*, Rockville, Md., 1992.

12.

Trademarks, The Who What Where and How of Black & Decker Trademarks, Black and Decker Manufacturing Co., Towson, Md., 1987.

13.

Chem. Week **130**(2), 28 (Jan. 13, 1982).

14.

United States Trademark Association, *The State Trademark Statutes*, New York, 1989. A looseleaf book (with supplements) containing the trademark laws of the 50 states.

15.

Section 44, *Trademark Act of 1946*, as amended, 15 U.S.C. § 1126.

General References

Bureau of National Affairs, *U.S. Patents Quarterly*, Washington, D.C. This weekly publication contains decisions of the courts and Patent and Trademark Office in patent, trademark, and copyright cases.

Compu-Mark Corporation, *Compu-Mark Directory*, North Quincy, N.J., 1991.

G. Hawley, *Condensed Chemical Dictionary*, Reinhold Co., New York, 1987.

International Bureau of the World Intellectual Property Organization, *Major Provisions of Trademark Legislation in Selected Countries*, Geneva, 1977.

S. Kane, *Trademark Law—A Practitioner's Guide*, Practicing Law Institute, New York, 1987.

J. T. McCarthy, *Trademarks and Unfair Competition*, Lawyers Cooperative Publishing Co., Rochester, N.Y., 1988.

Medical Economics Co., *Physicians' Desk Reference for Nonprescription Drugs*, Van Nostrand Reinhold, New York, 1991.

M. Windholz, ed., *Merck Index*, 11th ed., Merck & Co., Inc., Rahway, N.J., 1989.

Patent and Trademark Office, U.S. Department of Commerce, *Official Gazette*, Washington, D.C., published weekly. The trademark section, available separately, lists marks accepted, subject to opposition, with details of the application and all marks actually registered.

Patent and Trademark Office, U.S. Department of Commerce, *Trademark Rules of Practice of the Patent and Trademark Office with Forms and Statutes*, Washington, D.C., 1968. This publication can be purchased from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C.

Patent Office Society, *Journal of the Patent Office Society*, Washington, D.C., published monthly.

Thomas Register of American Manufacturers, Thomas Publishing Co., New York, 1992.

United States Trade Association (USTA), *The Trademark Handbook*, New York, 1990. The USTA has a number of publications on trademark-related subjects. A list of USTA publications is available from the USTA office, 6 East 45th Street, New York, N.Y. 10017.

United States Trademark Association, *The Trademark Reporter*, New York, published monthly.

Lile H. Deinard
Nicholas J. Stathis
Mary E. Rasenberger
White & Case

CORONARY THERAPEUTICS.

See CARDIOVASCULAR AGENTS.

CORROSION AND CORROSION CONTROL

Corrosion is the natural degradation of materials in the environment through electrochemical or chemical reaction. Traditionally, the definition of corrosion refers to the degradation of metals and has not included the degradation of nonmetals such as wood (rotting) or plastics (swelling or crazing), but increasingly, natural degradation of any engineering material is being regarded as corrosion. The vast majority of the technologically significant corrosion involves the deterioration of metallic materials, and only the corrosion of metallic materials is discussed here.

Electrochemical Nature of Corrosion

Ores are mined and are then refined in an energy intensive process to produce pure metals, which in turn are combined to make alloys (see METALLURGY; MINERAL RECOVERY AND PROCESSING). Corrosion occurs because of the tendency of these refined materials to return to a more thermodynamically stable state (1–4). The key reaction in corrosion is the oxidation or anodic dissolution of the metal to produce metal ions and electrons



The ions, M^{n+} , formed by this reaction at a rate, k_1 , may be carried into a bulk solution in contact with the metal, or may form insoluble salts or oxides. In order for this anodic reaction to proceed, a second reaction which uses the electrons produced, ie, a reduction reaction, must take place. This second reaction, the cathodic reaction, occurs at the same rate, ie, $I_c = I_a$, where I_c and I_a are the cathodic and anodic currents, respectively. The cathodic reaction, in most cases, is hydrogen evolution or oxygen reduction.

The three elements necessary for corrosion are an aggressive environment, an anodic and a cathodic reaction, and an electron conducting path between the anode and the cathode. Other factors such as a mechanical stress also play a role. The thermodynamic and kinetic aspects of corrosion determine, respectively, if corrosion can occur, and the rate at which it does occur.

Manifestations of Corrosion

The most common form of corrosion is uniform corrosion, in which the entire metal surface degrades at a near uniform rate (1–3). Often the surface is covered by the corrosion products. The rusting of iron (qv) in a humid atmosphere or the tarnishing of copper (qv) or silver alloys in sulfur-containing environments are examples (see also SILVER AND SILVER ALLOYS). High temperature, or dry, oxidation, is also usually uniform in character. Uniform corrosion, the most visible form of corrosion, is the least insidious because the weight lost by metal dissolution can be monitored and predicted.

An especially insidious type of corrosion is localized corrosion (1–3,5) which occurs at distinct sites on the surface of a metal while the remainder of the metal is either not attacked or attacked much more slowly. Localized corrosion is usually seen on metals that are passivated, ie, protected from corrosion by oxide films, and occurs as a result of the breakdown of the oxide film. Generally the oxide film breakdown requires the presence of an aggressive anion, the most common of which is chloride. Localized corrosion can cause considerable damage to a metal structure without the metal exhibiting any appreciable loss in weight. Localized corrosion occurs on a number of technologically important materials such as stainless steels, nickel-base alloys, aluminum, titanium, and copper (see ALUMINUM AND ALUMINUM ALLOYS; NICKEL AND NICKEL ALLOYS; STEEL; and TITANIUM AND TITANIUM ALLOYS).

Two types of localized corrosion are pitting and crevice corrosion. Pitting corrosion occurs on exposed metal surfaces, whereas crevice corrosion occurs within occluded areas on the surfaces of metals such as the areas under rivets or gaskets, or beneath silt or dirt deposits. Crevice corrosion is usually associated with stagnant conditions within the crevices. A common example of pitting corrosion is evident on household storm window frames made from aluminum alloys.

Another type of corrosion is dealloying which has also been called parting or selective leaching. Dealloying (1–3) is the preferential removal of one of the alloying elements from an alloy resulting in the enrichment of the other alloying element(s). Common examples are the loss of zinc from brasses (dezincification) and the loss of iron from cast irons (graphitization) (see COPPER ALLOYS).

Corrosion may also appear in the form of intergranular attack, ie, preferential attack of the grain boundary region (1–3). Often, intergranular attack is the result of stress corrosion, but it can also occur because the grain boundary and the grain have different corrosion tendencies, ie, different potentials. In the latter case, the grain boundary acts as the anode and the grain as the cathode. Intergranular corrosion can lead to a loss in strength or ductility of the metal.

Corrosion also occurs as a result of the conjoint action of physical processes and chemical or electrochemical reactions (1–3). The specific manifestation of corrosion is determined by the physical processes involved. Environmentally induced cracking (EIC) is the failure of a metal in a corrosive environment and under a mechanical stress. The observed cracking and subsequent failure would not occur from either the mechanical stress or the corrosive environment alone. Specific chemical agents cause particular metals to undergo EIC, and mechanical failure occurs below the normal strength (yield stress) of the metal. Examples are the failure of brasses in ammonia environments and stainless steels in chloride or caustic environments.

When a stress is cyclic rather than constant, the failure is termed corrosion fatigue. Fretting corrosion results from the relative motion of two bodies in contact, one or both being a metal. The motion is small such as a vibration. Erosion corrosion results from the action of a high velocity fluid impinging on a metal surface. Metals and alloys can also experience cracking in liquid metal environments. This form of corrosion is referred to as liquid metal cracking (LMC) (1,2,6,7). For example, mercury (qv) promotes cracking in highly stressed copper alloys and high strength aluminum alloys, and liquid copper promotes cracking of steels and stainless steels. Titanium and nickel alloys are also susceptible to LMC in specific environments.

Galvanic corrosion (1–3,8) occurs as a result of the electrical contact of different metals in an aggressive environment. The driving force is the electrode potential difference between the two metals. One metal acts principally as a cathode and the other metal as the anode. Galvanic corrosion can result from the presence of a second phase in a metal. An example is manganese sulfide [18820-29-6], MnS, inclusions in steels. Galvanic corrosion is also an important consideration for the environmental stability of metal matrix composites (qv) such as graphite reinforced aluminum. Galvanic corrosion can accelerate many of the other types of corrosion.

Microbiologically influenced corrosion, which results from the interaction of microorganisms and a metal, is receiving increased emphasis (1,3,9). The action of microorganisms is at least one of the reasons why natural seawater is more corrosive than either artificial seawater or sodium chloride solutions. Microorganisms attach to the surfaces of metals and can, for example, act as diffusion barriers; produce metabolites that enhance or initiate

corrosion; act as sinks or sources for species involved in cathodic reactions, such as oxygen and hydrogen; increase the pH at the surface as a result of photosynthesis; or decrease the pH by production of acid metabolites. A more detailed discussion of the various forms of corrosion may be found in the literature (1–3,6).

Origin of Corrosion

In the presence of oxygen and water the oxides of most metals are more thermodynamically stable than the elemental form of the metal. Therefore, with the exception of gold, the only metal which is thermodynamically stable in the presence of oxygen, there is always a thermodynamic driving force for corrosion of metals. Most metals, however, exhibit some tendency to passivate, ie, to form a protective oxide film on the surface which retards further corrosion.

The thermodynamics of electrochemical reactions can be understood by considering the standard electrode potential, E^0 , the potential of a reaction under standard conditions of temperature and pressure where all reactants and products are at unit activity. Table 1 lists a variety of standard electrode potentials. The standard potential is expressed relative to the standard hydrogen reference electrode potential in units of volts. A given reaction tends to proceed in the anodic direction, ie, toward the oxidation reaction, if the potential of the reaction is positive with respect to the standard potential. Conversely, a movement of the potential in the negative direction away from the standard potential encourages a cathodic or reduction reaction.

Table 1. Standard Electrode (Reduction) Potentials, 25°C^a

Electrode	E^0, V^{b}
$\text{Li}^+ + e^- \rightarrow \text{Li}$	−3.045
$\text{Zn}^{2+} + 2 e^- \rightarrow \text{Zn}$	−0.763
$\text{Fe}^{2+} + 2 e^- \rightarrow \text{Fe}$	−0.440
$\text{Cd}^{2+} + 2 e^- \rightarrow \text{Cd}$	−0.403
$\text{Co}^{2+} + 2 e^- \rightarrow \text{Co}$	−0.277
$\text{Ni}^{2+} + 2 e^- \rightarrow \text{Ni}$	−0.250
$\text{Sn}^{2+} + 2 e^- \rightarrow \text{Sn}$	−0.140
$\text{Pb}^{2+} + 2 e^- \rightarrow \text{Pb}$	0.126
$\text{Sn}^{4+} + 2 e^- \rightarrow \text{Sn}^{2+}$	0.15
$\text{Cu}^{2+} + e^- \rightarrow \text{Cu}^+$	0.158
$\text{Cu}^{2+} + 2 e^- \rightarrow \text{Cu}$	0.337
$\text{Fe}^{3+} + e^- \rightarrow \text{Fe}^{2+}$	0.771
$\text{Ag}^+ + e^- \rightarrow \text{Ag}$	0.799
$\text{Br}_2(\text{l}) + 2e^- \rightarrow 2\text{Br}$	1.065
$\text{Cl}_2(\text{g}) + 2e^- \rightarrow 2\text{Cl}$	1.358

^a Ref. 3.
^b Against standard hydrogen electrode (SHE) set at $E^0 = 0.000$ volts.

A piece of zinc placed in an aqueous 1 M Zn^{2+} solution attains a potential of approximately −0.763 V vs standard hydrogen. If Cu^{2+} ions are then introduced into the solution, the zinc begins to corrode because the zinc surface exists at a potential which encourages the cathodic reaction for copper, ie, the standard potential of the zinc reaction is negative to that of the copper and the Cu ions electroplate onto the zinc via the reaction $\text{Cu}^{2+} + 2 e^- \rightarrow \text{Cu}$. This coupling causes the zinc potential to shift, ie, to be polarized, in the positive direction and the reaction $\text{Zn} \rightarrow \text{Zn}^{2+} + 2 e^-$ then proceeds. The potentials of the two reactions are polarized toward each other, away from the respective standard potentials. The rates of these two reactions must be the same to preserve electroneutrality, and additionally, some zinc must be exposed to the solution. That is, if a continuous coating of copper were to be electroplated on the zinc surface such that none of the zinc were exposed, further oxidation would be prevented.

Corrosion occurs even if the two reactants involved are not at standard conditions. In this case the nonstandard equilibrium potential for each reaction, often referred to as the reversible potential, can be calculated from the Nemst equation. Additional information on thermodynamic aspects of corrosion can be found in Reference 10.

The larger the potential difference between the equilibrium potentials of the anodic and cathodic reactions involved in corrosion, the greater the driving force which is more traditionally represented as the change in Gibbs free energy, ΔG . The free-energy change for a reaction is related to the difference in the equilibrium potentials of the two reactions, cathodic minus anodic by

$$\Delta G = -nFE$$

(2)

where n is the number of electrons involved in the overall reaction, F is the Faraday constant, and E , the cell voltage or the difference in the equilibrium potential values of the cathode and the anode, respectively, given in volts.

The standard reduction potential of oxygen



is 1.229 V, a much more positive number than the metal reactions given in Table 1. Therefore, the free-energy change for a coupling of any of the metal reactions, considered as the anode, and the oxygen reaction, considered as the cathode, would be a large negative number; ie, the reaction is thermodynamically highly favored to proceed. Reduction of oxygen is the most common cathodic reaction in the corrosion of metals in neutral and alkaline media. In acidic media the common cathodic reaction is hydrogen evolution



The standard potential for this reaction is 0.000 V. Most of the metal reactions have standard potentials below this value.

Electrochemical Equilibrium Diagrams

The thermodynamic data pertinent to the corrosion of metals in aqueous media have been systematically assembled in a form that has become known as Pourbaix diagrams (11). The data include the potential and pH dependence of metal, metal oxide, and metal hydroxide reactions and, in some cases, complex ions. The potential and pH dependence of the hydrogen and oxygen reactions are also supplied because these are the common corrosion cathodic reactions. The Pourbaix diagram for the iron–water system is given as Figure 1.

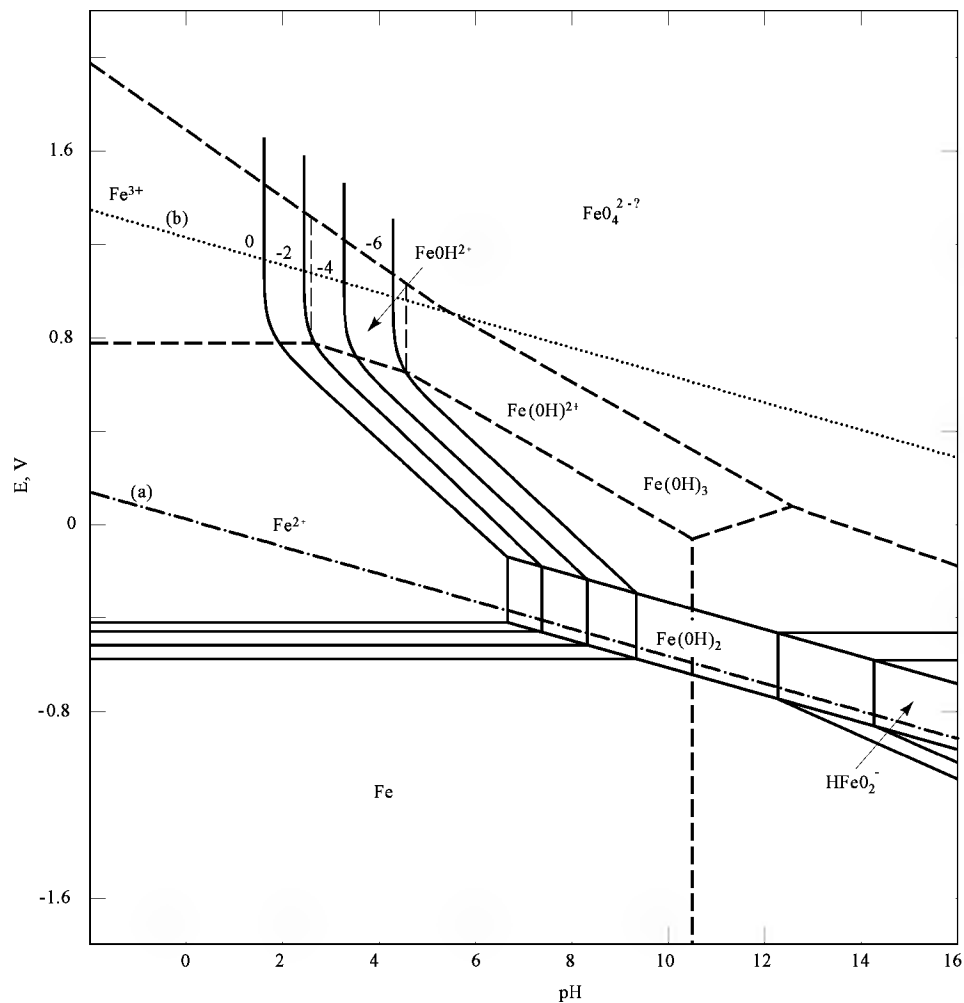


Fig. 1. Pourbaix diagram for the iron–water system at 25°C, considering as solid substances only Fe, Fe(OH)₂, and Fe(OH)₃, where (—), line a, represents the hydrogen reaction line; (· · ·), line b, the oxygen line (11). The values 0, –2, –4, –6 are molar concentrations of the ions involved in each reaction.

If the potential of a metal surface is moved below line a, the hydrogen reaction line, cathodic hydrogen evolution is favored on the surface. Similarly a potential below line b, the oxygen reaction line, favors the cathodic oxygen reduction reaction. A potential above the oxygen reaction line favors oxygen evolution by the anodic oxidation of water. In between these two lines is the region where water is thermodynamically stable.

The Pourbaix diagram is constructed from data for a variety of reactions. The expressions for these reactions can be: (1) potential dependent and pH independent, (2) pH dependent and potential independent, (3) both pH and potential dependent, or (4) both pH and potential independent. The dependence is determined by whether the reaction of interest is electrochemical or chemical and whether or not it involves the components of water. From this diagram it is possible to determine if there is a driving force for corrosion at any given potential–pH combination and what the corresponding cathodic reaction is likely to be. Regions where insoluble corrosion products form are indicated. These are regions where passivity is possible. Passivity generally results from the formation of an oxide film on a metal surface. The film dramatically reduces further corrosion of the metal. Regions of immunity, where there is no thermodynamic driving force for the metal to corrode, can also be determined. Figure 2 presents a simplified picture of the Pourbaix diagram of Figure 1 showing the regions of immunity, possible corrosion, and possible passivation. A complete collection of electrochemical equilibrium diagrams can be ordered from the National Association of Corrosion Engineers (NACE) (11).

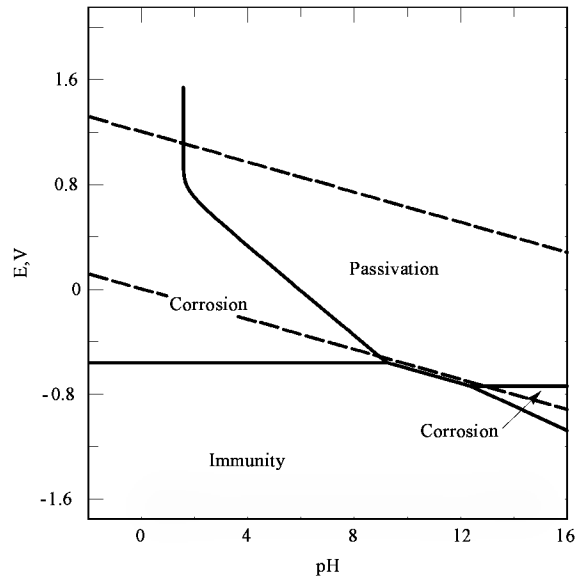


Fig. 2. The thermodynamic regions of corrosion, immunity, and passivation of iron in an iron–water system assuming passivation by a film of Fe₂O₃ (11).

Pourbaix diagrams are only thermodynamic predictions and yield no information about the kinetics of the reactions involved nor are the influences of other ionic species which may be present in the solution included. Complexing ions, particularly halides, can interfere with passivation and can influence

the position of the lines in a Pourbaix diagram. Further, in real corrosion situations the local chemistry in pits, crevices, and cracks differs from that of the bulk concentration and influences the actual electrochemical reactions at those sites. More comprehensive information on Pourbaix diagrams and the use of them in corrosion is available (1–3,10,11).

Kinetics of Electrochemical Reactions

Even in uniform corrosion, a corroding metal surface has numerous local anodes and cathodes. The sites of these local reactions may be fixed by microstructural features or may change as corrosion proceeds. The oxidation reaction at anodic sites on the metal surface can be represented as in equation 1. A corresponding reduction reaction must be occurring at cathodic sites. In acidic solutions this would likely be hydrogen evolution as shown in equation 4. The potentials of these two reactions would be moved toward each other, away from the respective equilibrium potentials, and the metal surface would assume an overall uniform potential. The dotted lines in Figure 3 illustrate the relationships of the individual electrochemical reaction kinetics as influenced by electrochemical potential. These lines also illustrate the interaction of the two reactions on the metal surface. Diagrams of this sort are referred to as Evans diagrams.

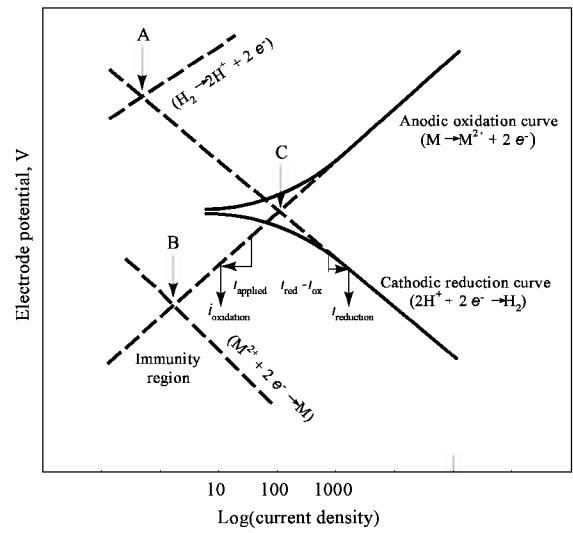


Fig. 3. Hypothetical Evans diagram and polarization curve for a metal corroding in an acidic solution, where point A represents the current density, i_0 , for the hydrogen electrode at equilibrium; point B, the exchange current density at the reversible or equilibrium potential, E_{rev} , for $M \rightleftharpoons M^{2+} + 2e^-$; and point C, the corrosion current density, i_{corr} , at the open-circuit corrosion potential, E_{corr} . See also discussion in text.

The two dashed lines in the upper left hand corner of the Evans diagram represent the electrochemical potential vs electrochemical reaction rate (expressed as current density) for the oxidation and the reduction form of the hydrogen reaction. At point A the two are equal, ie, at equilibrium, and the potential is therefore the equilibrium potential, E^0 , for the specific conditions involved. Note that the reaction kinetics are linear on these axes. The change in potential for each decade of log current density is referred to as the Tafel slope (12). Electrochemical reactions often exhibit this behavior and a common Tafel slope for the analysis of corrosion problems is 100 millivolts per decade of log current (1). A more detailed treatment of Tafel slopes can be found elsewhere (4,13,14).

The value of the current density where the oxidation and reduction forms of a reaction cross, ie, at point A or point B in Figure 3, is referred to as the exchange current density, i_0 . This is the rate of the two at equilibrium, and the greater the exchange current density the more kinetically favored is the reaction. The exchange current density for the reduction of hydrogen on platinized platinum is 10^{-3} A cm^2 , whereas on mercury it is 10^{-13} A cm^2 , indicating that platinum is a good catalyst for hydrogen evolution. The exchange current densities for most electrochemical reactions fall in between these two values. A thorough listing of exchange current densities for the hydrogen reaction on metal surfaces is available (15). Exchange current density and Tafel slope values for a variety of other electrochemical reactions of interest to corrosion can be found in the literature (16,17). The example in Figure 3 involves hydrogen evolution on a corroding surface and the exchange current density value for the hydrogen reaction on that surface must then be applied.

In the lower left part of the Evans diagram, the oxidation and reduction forms of some metal dissolution reactions are given. Assuming the data of this diagram are accurate, a prediction of corrosion rate and the potential of the corroding metal surface can be made. The potentials of the hydrogen and the metal reactions are pulled toward one another following the dashed lines, or Tafel slope, appropriate for the particular polarization; ie, the anodic line is followed for the metal dissolution reaction, the cathodic line followed for the hydrogen reaction. Point C where the two cross is the only potential where the rates of the anodic and the cathodic reactions are equal. This defines the corrosion potential, E_{corr} , and the corrosion current density, i_{corr} , from which the corrosion can be calculated by Faraday's Law. The other dotted lines were not considered because the rates at the corrosion potential are so low as to be insignificant.

The solid line in Figure 3 represents the potential vs the measured (or the applied) current density. Measured or applied current is the current actually measured in an external circuit; ie, the amount of external current that must be applied to the electrode in order to move the potential to each desired point. The corrosion potential and corrosion current density can also be determined from the potential vs measured current behavior, which is referred to as polarization curve rather than an Evans' diagram, by extrapolation of either or both the anodic or cathodic portion of the curve. This latter procedure does not require specific knowledge of the equilibrium potentials, exchange current densities, and Tafel slope values of the specific reactions involved. Thus Evans diagrams, constructed from information contained in the literature, and polarization curves, generated by experimentation, can be used to predict and analyze uniform and other forms of corrosion. Further treatment of these subjects can be found elsewhere (1–3,6,18).

Galvanic Corrosion and Cathodic Protection

Galvanic corrosion results when two or more dissimilar metals or alloys immersed in the same electrolyte are in electrical contact. This form of corrosion is one of the most common. The metal or alloy in a galvanic couple having the lower corrosion potential has its potential pulled in the positive direction by the coupling with the metal or alloy which has the higher corrosion potential, generally causing the one with the lower potential to experience accelerated corrosion. Conversely, the metal with the higher corrosion potential experiences a negative shift in potential as a result of the coupling causing it to corrode less and to support additional cathodic reaction. Figure 4 is a galvanic series for a variety of metals and alloys listing the corrosion potentials of these materials in seawater. Free corrosion potentials, not coupled or equilibrium potentials, are given and these allow a determination of which metal has the lower potential in a galvanic couple. Generally the larger the potential difference between components in a galvanic couple the higher the corrosion of the one having the lower potential. However, the kinetics of the particular reactions involved must be evaluated for an accurate determination of corrosion rate. This can be done using Evans diagrams or polarization curves (1,3).

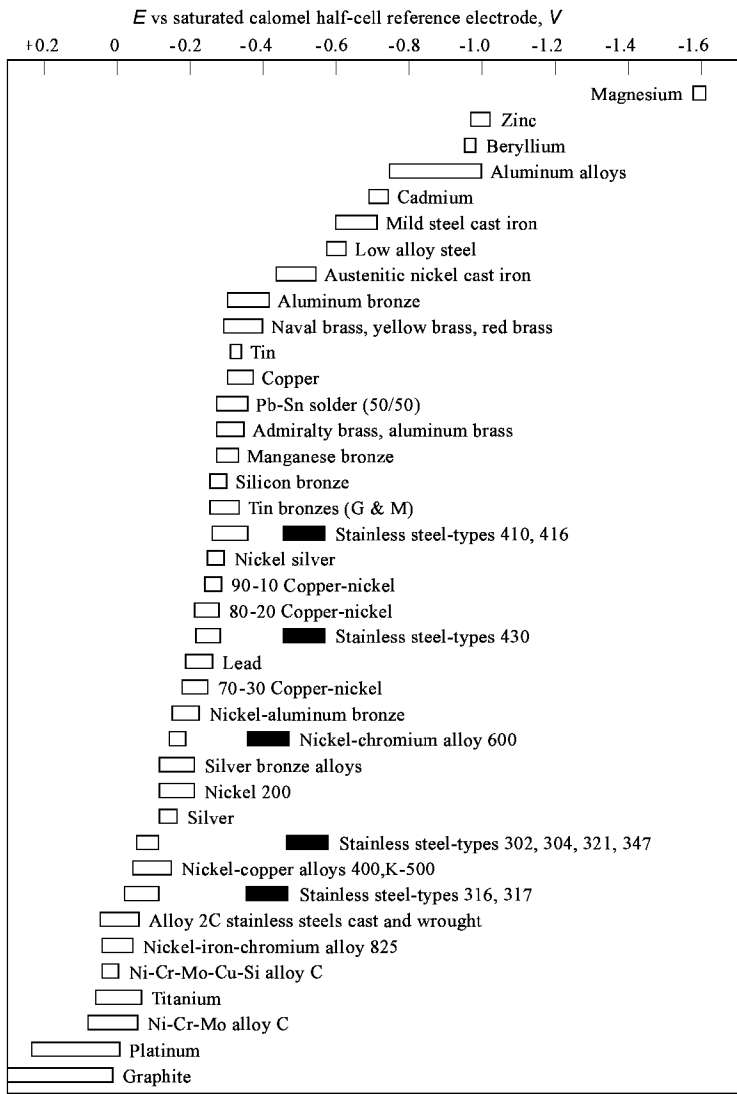


Fig. 4. Galvanic series in flowing seawater. Certain alloys may become more active in low velocity or poorly aerated seawater and the potentials exhibited under these conditions are indicated as darkened potential ranges (1).

Galvanic corrosion can be used to a corrosion advantage. If, for example, a metal such as zinc or magnesium, having a low corrosion potential in most environments, is coupled with a steel component the zinc or magnesium pulls the potential of the steel down causing the steel to corrode less. When a sacrificial metal or alloy, called a sacrificial anode, is attached to a structure having a higher corrosion potential to intentionally pull the potential of the higher potential metal down and thus decrease the corrosion rate, it is called cathodic protection. This method of corrosion mitigation is common for underground pipelines (qv), residential hot water heaters, and hulls and tanks (qv) of ships. The lowering of potential can also be achieved by the application of external electrical current, ie, impressed current cathodic protection. Electrons are pumped into the structure to lower the potential and thus discourage anodic (corrosion) reactions. A second electrode (anode) is needed for this procedure and this electrode is often a highly inert material such as platinum to prevent it from being corroded away. However, scrap iron or steel is sometimes used as the impressed current anode. This approach remains functional until the scrap metal anode corrodes away.

The lower potential metal in a galvanic couple does not always have its corrosion rate accelerated. For metals that form a passive film, coupling with another metal of higher potential can cause the film-forming metal's potential to shift from a value at which it corrodes to one at which it passivates and therefore corrodes less. When this is done intentionally the procedure is referred to as anodic protection, ie, achieving protection by intentionally shifting the potential in the positive direction. Anodic protection can also be brought about by adding oxidizers to the electrolyte.

The higher potential metal in a galvanic couple is not always rendered less corrodible. A passive metal can sometimes be pulled out of its passive region and into a more corrosive region. A metal can also have its potential pulled down to the point where hydrogen reaction is possible and this can lead to one of the forms of EIC called hydrogen induced cracking (HIC). Another example involves organic coatings (qv) on metal structures. These coatings can be damaged by having the potential of the structure shifted too far in the negative direction. This is one of the consequences of excessive cathodic protection.

Environmental Effects

The environment plays several roles in corrosion. It acts to complete the electrical circuit, ie, supplies the ionic conduction path; provide reactants for the cathodic process; remove soluble reaction products from the metal surface; and or destabilize or break down protective reaction products such as oxide films that are formed on the metal. Some important environmental factors include: the oxygen concentration; the pH of the electrolyte; the temperature; and the concentration of anions.

Reduction of oxygen is one of the predominant cathodic reactions contributing to corrosion. Awareness of the importance of the role of oxygen was developed in the 1920s (19). In classical drop experiments, the corrosion of iron or steel by drops of electrolytes was shown to depend on electrochemical action between the central relatively unaerated area, which becomes anodic and suffers attack, and the peripheral aerated portion, which becomes cathodic and remains unattacked. In 1945 the linear relationship between rate of iron corrosion and oxygen pressure from 0–2.5 MPa (0–25 atm) was shown (20).

The concentration dependence of iron corrosion in potassium chloride [7447-40-7], sodium chloride [7647-14-5], and lithium chloride [7447-44-8] solutions is shown in Figure 5 (21). In all three cases there is a maximum in corrosion rate. For NaCl this maximum is at approximately 0.5 N (about 3 wt %). Oxygen solubility decreases with increasing salt concentration, thus the lower corrosion rate at higher salt concentrations. The initial increase in the iron corrosion rate is related to the action of the chloride ion in concert with oxygen. The corrosion rate of iron reaches a maximum at ca 70°C. As for salt concentration, the increased rate of chemical reaction achieved with increased temperature is balanced by a decrease in oxygen solubility.

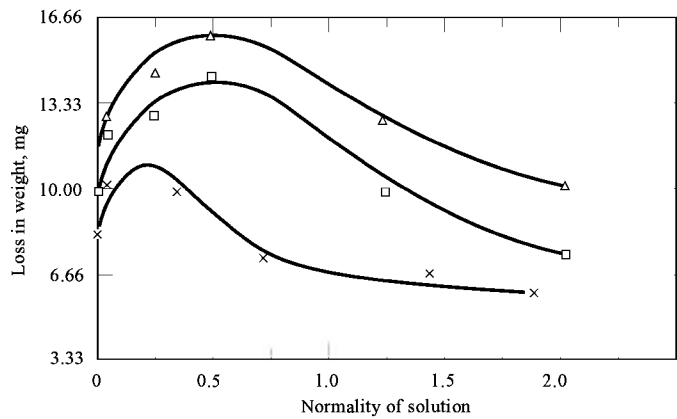


Fig. 5. Corrosion-concentration curves for alkali chlorides were ^ denotes LiCl; □, NaCl; and △, KCl (21).

The corrosion rate of iron in aerated water is also a function of pH and generally follows the pattern in Figure 6 (22). At 4–10 pH, the rate is controlled by the availability of oxygen. In more acidic solutions (lower pH) the corrosion rate is accelerated, and the reduction of hydrogen ion replaces the reduction of oxygen as the rate controlling cathodic reaction. Many metals follow approximately the same behavior with the exception of those metals that dissolve to form amphoteric ions. Zinc forms the zincate ion, ZnO_2^{2-} , which causes zinc to corrode excessively above a pH of 12 (see ZINC COMPOUNDS); whereas Al forms the aluminate ion [11098-82-1], AlO_2^- , which increases the dissolution rate above a pH of about 8 (see ALUMINUM COMPOUNDS, ALUMINATES).

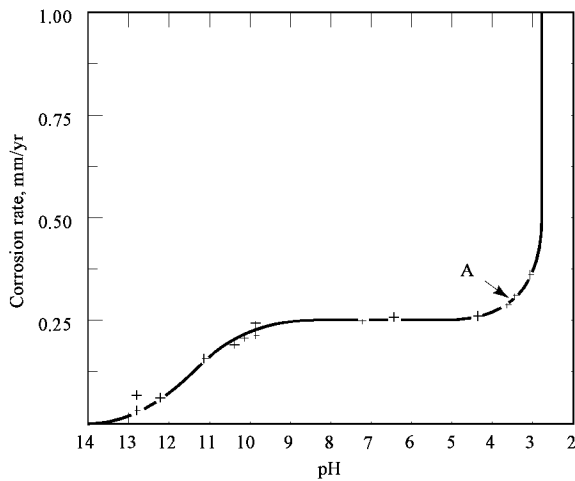


Fig. 6. Effect of pH on corrosion of iron in aerated water at room temperature (22). Point A is where H_2 evolution begins. To convert mm/yr to mils per year (mpy), multiply by 39.37.

Chlorides, which are ubiquitous in nature, play an important role in the corrosion of metals. Chlorides and other anions also play an important role in localized corrosion, ie, the breakdown of the insoluble protective reaction product films, eg, passive films, that prevent corrosion of the underlying metal. A variety of mechanisms attempting to explain the role of chloride in general and in localized corrosion have been proposed (23–25).

Very often the environment is reflected in the composition of corrosion products, eg, the composition of the green patina formed on copper roofs over a period of years. The determination of the chemical composition of this green patina was one of the first systematic corrosion studies ever made (see COPPER). The composition varied considerably depending on the location of the structure as shown in Table 2 (26,27).

Table 2. Composition of Green Patina on Copper from Different Locations^a

Location of structure	Age of structure, yr	Composition of green patina, %			
		CuCO_3	CuCl_2	Cu(OH)_2	CuSO_4
urban	30	14.6		9.6	49.8
rural	300	1.4		58.5	25.6
marine	13	12.8	26.7	52.5	2.5
urban–marine	38		4.6	61.5	29.7

^a Refs. 26 and 27.

Metallurgical Factors

The primary determining factor in corrosion behavior of metals and alloys is usually chemical composition. The amount of chromium in a stainless steel is a good example. There are, however, other metallurgical factors, such as crystallography, grain size and shape, grain heterogeneity, second phases, impurity inclusions, and residual stress, that influence corrosion. The technologically important structural materials are polycrystalline aggregates. Each individual crystal is referred to as a grain. Grain orientation can affect corrosion resistance as evidenced by metallographic etching rates and pitting behavior (23). Grain shape may likewise vary greatly depending on the alloy and processing history.

Alloys, particularly in the as-cast condition, generally exhibit chemical inhomogeneity such that there is segregation of alloying elements and impurities to the grain boundary regions. These heterogeneities, which can also develop during subsequent processing such as welding (qv) or heat treatment, can produce different electrochemical characteristics at the grain boundary relative to the grain interior and can lead to intergranular corrosion. This problem can be of great practical importance, especially to wrought stainless steels and nickel alloys. Second phases, such as ferrite grains in an otherwise austenitic stainless steel and beta grains in an otherwise alpha brass, can be of considerable practical importance in some alloy systems and some forms of corrosion. Residual stresses from cold-working or other sources can lead to increased corrosion rates (23,24,28,29) and are also important in

stress-corrosion cracking.

A complete discussion of the corrosion behavior of alloy systems and the influence of metallurgical factors on each is available (30). Some of these factors in a few technologically important alloy systems are discussed here.

Stainless Steels. The most common and serious metallurgical factor affecting the corrosion resistance of stainless steels is termed sensitization. This condition is caused by the precipitation of chromium-rich carbides (qv) at the grain boundaries, giving rise to chromium depleted grain boundary areas. These areas are anodic to the grain interior and tend to dissolve thereby causing intergranular corrosion. Sensitization also makes stainless steel more prone to stress-corrosion cracking in chloride environments at elevated temperature. The precipitation responsible for sensitization occurs in the range of roughly 425–900°C and may take place during relatively slow cooling from forging, welding, or during annealing when, for example, a stainless steel component attached to carbon steel pressure vessel must be stress-relieved thermally.

There are several measures available to mitigate sensitization. Low carbon grades such as AISI 304L and 316L are available which have much less tendency toward sensitization. Additionally, stabilized grades such as AISI 321 and 347 that contain titanium and niobium, respectively, to tie up the carbon are available. Sensitization can be detected chemically, eg, using ASTM A262 methods, employing test specimens (coupons) that have been processed with the structure or component of interest.

Sulfide inclusions have been identified as the origins of pits and stress-corrosion cracks in stainless steel structures under some conditions (see STEEL).

Copper Alloys. Various names are associated with brasses of a particular zinc content: 40% Zn–Cu alloys are referred to as muntz metal; 30% Zn–Cu alloys are referred to as yellow brass; and 15% Zn–Cu alloys are termed red brass. Brasses are susceptible to dealloying in the form of dezincification, ie, the preferential loss of zinc from the alloy. Brasses having zinc concentrations of 15% or greater are prone to dezincification and dezincification is generally more severe in brasses that have two metallurgical phases. For the two-phase alloys, greater than approximately 38 wt % Zn, dezincification usually starts in the high zinc content β -phase and is followed by dezincification of the lower zinc α -phase (31). Naval brass and admiralty brass are similar in composition to muntz metal and yellow brass, respectively, but have 1 wt % tin added for improved resistance to dezincification. Red brass is relatively resistant to dezincification but is more susceptible to impingement attack than, for example, yellow brass.

Conditions that favor dezincification include stagnant solutions, especially acidic ones, high temperatures, and porous scale formation (2). Additions of small amounts of arsenic, antimony, or phosphorus can increase the resistance to dezincification. These elements are, however, not entirely effective in preventing the dezincification of the two-phase (α – β) brasses because dezincification of the β -phase is not prevented (31). Another area of corrosion concern involves applied or residual stresses from fabrication that can lead to EIC of brasses in the form of stress-corrosion cracking.

Aluminum Alloys. Copper, silicon, magnesium, zinc, and manganese are some common alloying additions to aluminum. Most of the alloying elements are added to aluminum to produce alloys having improved mechanical properties. However, the strengthening phases which result from the alloying can disrupt the passive oxide layer on aluminum and lead to localized corrosion in the form of pitting attack. Also, these strengthening phases can result in intergranular corrosion.

Intergranular corrosion in aluminum alloys is caused by a potential difference, ie, galvanic couple, between the grain and grain boundary. The formation of the anodic area in the couple varies depending on the alloying additions (31). For example, in alloys containing copper, the anodic area is formed by the depletion of copper adjacent to CuAl_2 [12004-15-8] precipitates which form preferentially along the grain boundaries (2). In alloys containing magnesium, Mg_2Al_3 [12004-68-1] can form along the grain boundaries and this compound is the anodic constituent (31). Stress-corrosion cracking is characteristically intergranular for aluminum alloys. Wrought high strength aluminum alloys, whether the products are rolled, forged, or extruded, tend to be highly textured because of the manner in which the secondary phases or inclusions are strung out. This processing causes the grains of the primary α -phase to be flattened and elongated, as is shown schematically in Figure 7 (31,32). Whereas this metallurgical texture is of comparatively minor consequence to most forms of corrosion, it can promote exfoliation corrosion, intergranular corrosion leading to the leafing-off of uncorroded grain-bodies. Further, the texture is important to the stress-corrosion cracking of high strength aluminum alloys (see ALUMINUM AND ALUMINUM ALLOYS).

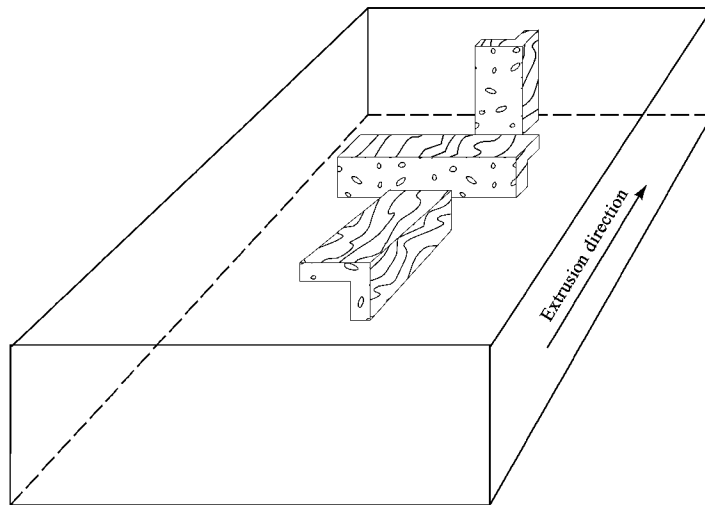


Fig. 7. Texture in a wrought high strength aluminum product (29). These elongated, flattened grains are produced by rolling, extruding, or forging.

Nickel Alloys. Nickel-based alloys have great technological importance in the development of high strength, high temperature alloys (qv), in part because of the occurrence of unique intermetallic phases (31). However, the precipitation of certain intermetallic phases can decrease the resistance to corrosion because this can deplete the matrix of alloying elements. For example, the nickel-base alloys containing chromium for passivity, which are of great importance to the chemical industry, can become sensitized in a manner similar to the austenitic stainless steels and thus become vulnerable to intergranular corrosion.

Environmentally Induced Cracking

Environmentally induced cracking (EIC) is a brittle fracture process caused by the conjoint action of a mechanical stress and a corrosive environment. There are several different types of EIC: stress corrosion cracking (SCC), corrosion fatigue cracking (CFC), and hydrogen induced cracking (HIC). In SCC and HIC the stress may be normal stress experienced in operation or residual stress from welding, heat treatment, and or cold work. CFC results from cyclic stresses. In any of these the corrosion rates are generally very low and the stress level required for the cracking is generally well below the normal yield stress of the particular alloy involved. The concentration of the chemical species responsible for the cracking need not be high. EIC failures are often catastrophic such as the Pt. Pleasant, West Virginia, bridge collapse that killed 46 people in 1967 (33). Cracking in steam pipes, underground pipelines, and aircraft have also been attributed to EIC.

The chemical species that can lead to EIC in each alloy system are fairly well known although the exact mechanisms of crack initiation and propagation are not thoroughly understood. The species that promote EIC in one alloy system do not necessarily promote EIC in others. A discussion of EIC and the chemical species that promote it can be found in the literature (34).

The cracking in EIC can proceed either intergranularly, ie, between the grains, or transgranularly, ie, across the grains. Small changes in the environment can cause a shift from one type to the other as can relatively small changes in the composition of the alloy. Likewise the crack may be single and unbranched or have multiple origins and also be branched. Thus neither intergranular or transgranular cracking is indicative of or excludes EIC as the cause of a service crack. More detailed information regarding EIC is available (1,3,6,30).

Copper Alloys. Copper alloys under an applied or residual stress are susceptible to stress-corrosion cracking in environments containing ammonia (qv) or ammonium compounds (qv). In addition to ammonia, water and oxygen, or another cathodic reactant, are required to cause SCC. Ammonia and ammonium compounds are sometimes found in the atmosphere, in cleaning compounds, in fertilizers (qv), or in chemicals used for water treatment. Brasses containing less than 15% zinc are highly resistant to stress-corrosion cracking (SCC), whereas brasses containing 20 to 40% zinc are highly susceptible and this susceptibility increases only slightly as zinc concentration increases from 20 to 40% (30). SCC is usually intergranular but can occur transgranular under certain conditions (2,31).

Trace quantities of nitrogen oxides may also cause SCC. These failures are probably the result of the nitrogen oxides being converted to ammonium salts on the brass surface. The premature failure of yellow brass brackets in the humidifier chamber of an air-conditioning system was traced to this latter cause (2). In another case, SCC of 12% Ni–23% Zn–Cu alloy, known as nickel brass, parts of the Central Office Telephone equipment in Los Angeles occurred within two years of installation for similar reasons. The source of nitrogen was the air in the Los Angeles area which has high concentrations of nitrogen oxides and suspended nitrates (see AIR POLLUTION). The nitrates settle as dust on the brass parts (2). The susceptibility to SCC can be minimized by: (1) proper alloy selection; (2) thermal stress relief; (3) avoiding contact with ammonia and ammonia compounds; and (4) using an inhibitor (2,31).

Aluminum Alloys. Both the 2000 and the 7000 series aluminum alloys have experienced significant SCC in service. These high strength alloys in the wrought form are highly textured (Fig. 7), and the SCC behavior differs greatly according to the direction of tensile stress with respect to the texture directions. The alloys are most vulnerable to SCC if stressed parallel to the short transverse grain direction, ie, parallel to the thinnest dimension of the grain; most resistant if stressed only parallel to the longest grain dimension; and show intermediate susceptibility if stressed parallel to the long transverse direction. In practice, it is the short transverse-direction stresses that cause SCC problems. Hence prudent practice is to avoid designs in which high sustained stresses are imposed across the short transverse direction. For example, one avoids an interference-fit fastener, or a taper pin fastener oriented to stress a part across the vulnerable texture unless the alloy is inherently of low susceptibility to SCC. Table 3 summarizes alloys and tempers in various categories of susceptibility as judged from specimens having short transverse, ie, most vulnerable, orientations.

Table 3. Susceptibility Classification of Commercial Wrought Aluminum Alloys in Plate Form—Short Transverse Orientation^a

Aqueous susceptibility	Aluminum alloy	Temper ^b
very low	1100	all
	3003, 3004, 3005	all
	5000, 5050, 5052, 5154	all
	5454, 6063	
	5086	O, H32, H34
	6061, 6262	O, T6
	Alclad: 2014, 2219, 6061, 7075	all
low	2219	T6, T8
	5086	H36
	5083, 5456	controlled
	6061	T4
	6161, 5351	all
	6066, 6070, 6071	T6
	2021	T8
	7049, 7050, 7075	T73
	2024, 2124	T8
	7050, 7175	T736
moderate	7049, 7075, 7178	T6
	2024, 2219	T3, T4
	2014, 7075, 7079, 7178	T6
appreciable	5083, 5086, 5456	sensitized
	7005, 7039	T5, T6

^a Ref. 35.

^b Standard mill designation.

In addition to the possibility of selecting alloys having minimum SCC susceptibility while retaining other properties as needed, there are other steps possible to reduce the SCC probability: (1) avoid designs that permit water to accumulate; (2) avoid conditions in which salts, especially chlorides, can concentrate; and (3) where available and otherwise acceptable, use a clad alloy (2) (see METAL SURFACE TREATMENTS).

High Strength Steels. Steels that owe their strength to heat treatment, whether martensitic, precipitation hardened, or maraging, and whether stainless or not, are susceptible to SCC in aqueous environments, including water vapor. The primary factor in determining the degree of SCC susceptibility of a given steel is its strength. There is no sharply defined threshold strength that defines a threshold susceptibility, but above 1200 MPa (174,000 psi) yield strength the problem becomes of increasing concern, until at 1400 MPa (ca 200,000 psi) susceptibility is generally very high. This does not mean that steels cannot be used at such strength levels or even higher, but for steels to be used at these strengths, great care must be exercised in design such that any tensile or bending stresses are small, or that moisture is excluded from the surface of the steel.

High strength steels are also susceptible to HIC. Cadmium electroplating is a useful protection measure for steels, but hydrogen either must not be codeposited with cadmium, or if it is codeposited, as in the cyanide plating bath, hydrogen must be safely redistributed by thermal treatment before the high strength steel component is stressed. Otherwise the steel may experience HIC.

Stainless Steels. Austenitic stainless steels undergo SCC when stressed in hot aqueous environments containing chloride ion. The oxygen level of the environment can be important, probably through its effect in establishing the electrode potential of the steel. The total matrix of combinations of stress level, chloride ion and oxygen concentrations, and temperature that cause SCC has not been worked out. At about yield strength stress and ca 290°C, 1 ppm chloride and 1 ppm dissolved oxygen are about the minimum levels that initiate SCC. At room temperature, chloride SCC seldom occurs in austenitic stainless steels except when the steels are heavily sensitized. When these steels are not heavily sensitized, there is not usually SCC except at elevated temperatures. Whereas there is no well-defined threshold temperature for vulnerability to SCC in stainless steels, above about 60°C, it becomes of increasing concern.

The relative susceptibilities of the common grades of austenitic stainless steel to SCC in chloride do not differ greatly, although the steels that can and do become sensitized are decidedly inferior in this condition. The high purity ferritic grades of stainless steel offer appreciable improvement in SCC resistance as compared to the austenitic grades, but they are not immune to cracking. Care must also be exercised to avoid the ductile-to-brittle transition that most steels undergo as the temperature is decreased and potential embrittlement by σ-phase formation in elevated temperature service. The standard methods for avoiding chloride SCC in austenitic stainless steels include: avoidance of fabrication stresses; minimizing chloride ion level; and minimizing oxygen concentration in the environment. Additionally, where feasible, it is possible to mitigate against SCC in these alloys by cathodic protection, such as through coupling to iron.

Corrosion-Resistant Materials

Alloys having varying degrees of corrosion resistance have been developed in response to various environmental needs. At the lower end of the alloying scale are the low alloy steels. These are iron-base alloys containing from 0.5–3.0 wt % Ni, Cr, Mo, or Cu and controlled amounts of P, N, and S. The exact composition varies with the manufacturer. The corrosion resistance of the alloy is based on the protective nature of the surface film, which in turn is based on the physical and chemical properties of the oxide film. As a rule, this alloying reduces the rate of corrosion by 50% over the first few years of atmosphere exposure. Low alloy steels have been used outdoors with protection.

The stainless steels contain appreciable amounts of Cr, Ni, or both. The straight chrome steels, types 410, 416, and 430, contain about 12, 13, and 16 wt % Cr respectively. The chrome–nickel steels include type 301 (18 wt % Cr and 9 wt % Ni), type 304 (19 wt % Cr and 10 wt % Ni), and type 316 (19 wt % Cr and 12 wt % Ni). Additionally, type 316 contains 2–3 wt % Mo which greatly improves resistance to crevice corrosion in seawater as well as general corrosion resistance. All of the stainless steels offer exceptional improvement in atmospheric conditions. The corrosion resistance results from the formation of a passive film and, for this reason, these materials are susceptible to pitting corrosion and to crevice corrosion. For example, type 304 stainless has very good resistance to moving seawater but does pit in stagnant seawater.

Several copper alloys are exceptionally resistant to certain atmospheres. The copper–nickel alloys, 90% Cu–10% Ni and 70% Cu–30% Ni, have very good resistance to corrosion in seawater, if the iron content is properly controlled, ie, from 0.5 to 1% Fe. Because of the high copper content, these alloys also resist microbiological fouling as well as corrosion in most aqueous environments. Monel alloy (66.5 wt % Ni–31.5 wt % Cu) is widely used in marine service where strength is also required. Several copper alloys, silicon bronzes, aluminum bronzes, and manganese bronzes are resistant to severe atmospheric corrosion conditions, and because of their high strength are used as bolts, clamps, or load carrying parts in outdoor environments. Specific alloys have been found suitable for very corrosive environments. For example, Carpenter 20, which is 20 wt % Cr–29 wt % Ni–2.5 wt % Mo–3.5 wt % Cu, has outstanding resistance to concentrated sulfuric acid.

For most environments quantitative studies have been reported describing the corrosion rate of various materials including a number of corrosion-resistant alloys (30). For example, Table 4 gives weight losses suffered by corrosion-resistant alloys in a solution of 28% phosphoric acid [7664-38-2], 20–22% sulfuric acid [7664-93-9], and 1–15% fluoride (36).

Table 4. Plant Corrosion Test in Sulfuric Acid Dilution with Recirculated Phosphoric Acid^a

Material	Corrosion rate, μm yr	Concentration cell depth, μm
Carpenter stainless no. 20Cb	28	
Aloyco 20	38	
Incoloy alloy 825	69	127
Hastelloy alloy C	71	
Illium "G"	76	
Inconel alloy 718 ^b	81 ^c	
Worthite	109	76
Incoloy alloy 901 ^d	155	127
Illium "R"	175, 363	
Monel alloy K-500 ^e	787	f
Monel alloy 400	965	
Hastelloy alloy B	1980	
stainless type 317	>3800	
stainless type 316	>4600	

^a Phosphoric acid (wet-process) 28% (20% P₂O₅), sulfuric acid 20–22%, fluoride approx 1–1.5%, probably as hydrofluosilicic acid; temperature 82–110°C, average 93°C; and duration of test 42 days, moderate aeration, agitation by convection only.

^b Composition 52.5% Ni, 18.6% Cr, 18.5% Fe, 5.0% Nb, and 3.1% Mo.

^c Pitted to a maximum depth of 127 μm.

^d Composition 42.7% Ni, 34.0% Fe, 13.5% Cr, 6.2% Mo, and 2.5% Ti.

^e Composition 65.0% Ni, 29.5% Cu, 2.8% Al, and 1.0% Fe.

^f Pitted to a maximum depth of 76 μm.

Inhibitors

Corrosion inhibitors are substances which slow down or prevent corrosion when added to an environment in which a metal usually corrodes. Corrosion inhibitors are usually added to a system in small amounts either continuously or intermittently. The effectiveness of corrosion inhibitors is partly dependent on the metals or alloys to be protected as well as the severity of the environment. For example, the main factors which must be considered before application of a corrosion inhibitor to an aqueous system are the compatibility of the inhibitor and the metal(s), the salt concentration, the pH, the dissolved oxygen concentration, and the concentration of interfering species such as chlorides or metal cations. In addition, many inhibitors, most notably chromates, are toxic and environmental regulations limit use. Attention is now being given to the development of more environmentally compatible inhibitors (37).

Inhibitors act and are classified in a variety of ways (1,3,37,38). The classifications used herein closely follow the discussion in Reference 37. Types of inhibitors include: (1) anodic, (2) cathodic, (3) organic, (4) precipitation, and (5) vapor-phase inhibitors.

Anodic Inhibitors. Passivating or anodic inhibitors produce a large positive shift in the corrosion potential of a metal. There are two classes of anodic inhibitors which are used for metals and alloys where the anodic shift in potential promotes passivation, ie, anodic protection. The first class includes oxidizing anions that can passivate a metal in the absence of oxygen. Chromate is a classical example of an oxidizing anodic inhibitor for the passivation of steels.

The second class of anodic inhibitors contains ions which need oxygen to passivate a metal. Tungstate and molybdate, for example, require the presence of oxygen to passivate a steel. The concentration of the anodic inhibitor is critical for corrosion protection. Insufficient concentrations can lead to pitting corrosion or an increase in the corrosion rate. The use of anodic inhibitors is more difficult at higher salt concentrations, higher temperatures, lower pH values, and in some cases, at lower oxygen concentrations (37).

Cathodic Inhibitors. Cathodic inhibitors act to retard or poison the cathodic reaction or selectively precipitate onto cathodic areas producing diffusion barriers to cathodic reactants, thereby reducing the rate of the cathodic reaction. There are three types of cathodic inhibitors: (1) hydrogen poisons, (2) oxygen scavengers, and (3) cathodic precipitates. Hydrogen poisons are chemical species such as arsenic or antimony that retard the hydrogen reduction reaction. Because the hydrogen poison slows the cathodic reaction, and because the cathodic and anodic reactions must proceed at the same rate, the whole of the corrosion process is slowed. A potentially serious drawback upon use of hydrogen poisons is that the hydrogen on the surface can be

more easily absorbed into the metal or alloy and can lead to HIC in susceptible materials.

Oxygen scavengers prevent corrosion by tying up the oxygen in solution, thereby making it unavailable for the cathodic reaction. The most common oxygen scavengers used in water at ambient temperatures are sulfur dioxide [7446-09-5] and sodium sulfate [7757-82-6]. Cathodic precipitate inhibitors such as calcium carbonate [471-34-1] or magnesium carbonate [546-93-0] precipitate onto the cathodic areas producing a film that reduces the cathodic activity (37).

Organic Inhibitors. Generally, organic inhibitors adsorb on the entire metal surface and impede corrosion reactions (37). Organic inhibitors consist of broad classes of organic compounds. For example, aliphatic organic amines (qv) adsorb by the surface-active–NH₂ group which forms a chemisorptive bond with the metal surface. The hydrocarbon tails orient away from the interface toward the solution, so that further protection is provided by the formation of a hydrophobic network which excludes water and aggressive ions from the metal surface (38). Organic inhibitors influence both anodic and cathodic reactions to varying degrees depending on the potential, the chemical structure of the inhibitor, and the size of the inhibitor molecule. Soluble organic inhibitors produce a protective layer that is only a few molecules thick, whereas insoluble inhibitors added as dispersions can build a film to a thickness of several hundredths of a centimeter (37).

Precipitation and Vapor-Phase Inhibitors. Precipitation inhibitors are film-forming compounds that produce barrier films over the entire surface. Phosphates and silicates, which are the most common, do not provide the degree of protection afforded by chromate inhibitors, but are useful in situations where nontoxic additives are required. Two main drawbacks to the use of phosphates and silicates are the dependence on the water composition and the control required to achieve maximum inhibition (37,38).

Vapor-phase inhibitors are volatile compounds that adsorb onto metal surfaces, and retard or prevent corrosion by a variety of mechanisms (37). Inhibitors such as dicyclohexamine nitrate [3882-06-02] can protect a variety of metals such as steel, aluminum, and tinplate. A number of vapor-phase inhibitors are commercially available as powders or tablets. However, vapor-phase inhibitors attack nonferrous metals to varying degrees, thus the manufacturers' recommendations should be checked before application. The system to be protected must be closed to maintain the volatile compound, but objects as large as the interior of an ocean-going tanker have been treated by this technique.

Coatings for Corrosion Prevention

Coatings (qv) are applied to metal substrates to prevent corrosion. Generally the coating protects the metal by imposing a physical barrier between the metal substrate and the environment. However, the coating can also act to provide corrosion protected by cathodic protection or by serving as holding reservoirs for inhibitors. Coatings may be divided into organic, inorganic, and metallic coatings (1–3,39,40). For an in-depth coverage of this topic see References 1–3 (see also COATINGS, MARINE; METAL SURFACE TREATMENTS).

Coating Types. Organic coatings afford protection by providing a physical barrier and are often used as holding reservoirs for corrosion inhibitors. Organic coatings include paints, resins, lacquers, and varnishes (see also BARRIER POLYMERS; PAINT). These coatings probably protect more metal on a tonnage basis than any other means of corrosion protection. As a rule, however, organic coatings should not be used in environments that would rapidly corrode the metal if the coating were compromised. For example, paint would not be used to protect the inside of a tank car used for shipping hydrochloric acid because one small coating defect would result in the rapid perforation of the tank car wall. The effective application of an organic coating requires: (1) proper surface preparation of the metal substrate, (2) selection of the proper primer, and (3) selection of the appropriate top coat(s). Poor performance of organic coatings most often results from improper surface preparation or poor application (1). Maintenance programs involving periodic inspection and repair are necessary for reliable corrosion protection by organic coatings.

Inorganic coatings are also used to provide a barrier between the environment and the metal (1,2,39). Inorganic coatings include chemical conversion coatings, glass (qv) linings, enamels (see ENAMELS, PORCELAIN OR VITREOUS), and cement (qv). Chemical conversion coatings are produced by intentionally corroding the metal surface in a controlled manner. This is done so as to produce an adherent corrosion product that protects the metal from further corrosion. Anodization of aluminum produces a protective aluminum oxide film on the aluminum metal. Other examples of chemical conversion coatings include phosphatizing for the protection of automobile bodies and chromating for the protection of zinc and magnesium. Porcelain enamel coatings which are inert in water and resistant to most weather are routinely applied to steel, cast iron, and aluminum. They are commonly seen on appliances and plumbing fixtures. Glass-lined metals are used in process industries where there is concern over corrosion or contamination of the product. Glass-lined steels are used in the pharmaceutical industry, breweries, and food plants. Portland cement coatings have been used to protect steel and cast-iron water pipes and have an excellent performance record.

In some cases, metallic coatings, in addition to providing a barrier between the metal substrate and the environment, provide cathodic protection when the coating is compromised (1,3,40). Metallic coatings (qv) are produced using techniques such as hot dipping, electroplating, cladding, flamespraying, chemical vapor deposition, or by surface modification using directed energy (laser or ion) beams.

Coating Techniques. Hot dipping is one of the oldest methods used for coating metals. In hot dipping, coatings are applied by dipping the metal in a molten bath commonly of zinc, tin, lead, or aluminum. The best known hot dipping procedure is that of coating steel with zinc to produce galvanized steel. In addition to providing a barrier between the steel substrate and the environment, the zinc acts as a sacrificial anode and provides cathodic protection to the underlying steel when the coating is breached.

Electroplating (qv) consists of immersing a metal in a plating bath and electrochemically plating the solution species onto the metal substrate. Additives to the plating bath can improve coating properties such as grain size, strength, uniformity, and brightness. The electroplate can be a single metal, an alloy, or a sequence of layers of different composition. Coating thicknesses can range from a few hundredths of a micrometer to upward of 500 μm. Zinc, nickel, tin, and cadmium are the most commonly plated materials on a tonnage basis. Electroplated tin is used as a protective coating for food cans (see FOOD PACKAGING). The electroplated tin provides a physical barrier and galvanic protection if the coating is compromised. However, the tin coating cannot provide galvanic protection in the presence of dissolved oxygen so that food should not remain in tin plated cans after opening.

The ability to modify metal surfaces using directed energy beams is providing a new flexibility in attempting to improve the corrosion resistance of metals. Techniques such as ion implantation (qv), ion beam mixing, and ion beam assisted deposition (IBAD), as well as laser-surface alloying and processing have been used to improve the corrosion behavior of metals (see LASERS). These techniques are versatile and have been used to modify the cathodic or anodic reactions, improve the nature of passive films, and produce a barrier coating (40). Ion implantation has been used to improve the corrosion resistance of titanium in hot acids and aluminum and steels in aqueous chloride environments (40).

BIBLIOGRAPHY

"Corrosion Inhibitors," in *ECT* 2nd ed., Vol. 6, pp. 317–346, by C. C. Nathan; "Corrosion and Corrosion Inhibitors," in *ECT* 3rd ed., Vol. 7, pp. 113–142, by R. T. Foley and B. F. Brown, The American University.

1. M. G. Fontana, *Corrosion Engineering*, 3rd ed., McGraw-Hill Book Co., Inc., New York, 1986.
2. H. H. Uhlig and R. W. Revie, *Corrosion and Corrosion Control*, 3rd ed., John Wiley & Sons, Inc., New York, 1985.
3. D. A. Jones, *Principles and Prevention of Corrosion*, Macmillan Publishing Co., New York, 1992.
4. E. Gileadi, E. Kirowa-Eisner, and J. Penciner, *Interfacial Electrochemistry*, Addison-Wesley Publishing Co., Reading, Mass., 1975.
5. Z. Szklarska-Smialowska, *Pitting Corrosion of Metals*, National Association of Corrosion Engineers, Houston, Tex., 1986.
6. L. S. Van Delinder, ed., *Corrosion Basics: An Introduction*, National Association of Corrosion Engineers, Houston, Tex., 1984.
7. H. L. Logan, in Ref. 6, p. 121.
8. J. W. Oldfield, in H. P. Hack, ed., *Galvanic Corrosion*, American Society for Testing and Materials, ASTM-STP 978, Philadelphia, Pa., 1988, p. 5.
9. S. C. Dexter, *Biofouling*, in press, 1991.
10. M. Pourbaix, *Lectures on Electrochemical Corrosion*, Plenum Press, New York-London, 1973.
11. M. Pourbaix, *Atlas of Electrochemical Equilibria in Aqueous Solutions*, NACE, Houston, Tex., 1974.

12. J. Tafel, *Z. Physik. Chem. (Leipzig)* **50**, 641 (1905).

13. J. O'M. Bockris and A. K. N. Reddy, *Modern Electrochemistry*, Vol. 2, Plenum Press, New York, 1970.

14. A. J. Bard and L. R. Faulkner, *Electrochemical Methods*, John Wiley & Sons, Inc., New York, 1980.

15. H. Kita, *J. Electrochem. Soc.* **113**, 1905 (1966).

16. Ref. 2, p. 49.

17. Ref. 3, pp. 98 and 99.

18. R. Baboian, ed., *Electrochemical Techniques in Corrosion Engineering*, NACE, Houston, Tex., 1986.

19. U. R. Evans, *J. Soc. Chem. Ind.* **43**, 315-22T (1924).

20. W. H. J. Vernon, *J. Sci. Inst.* **22**, 226 (1945).

21. C. W. Borgmann, *Ind. Eng. Chem.* **29**, 814 (1937).

22. W. Whitman, R. Russell, and V. Atteri, *Ind. Eng. Chem.* **16**, 665 (1924).

23. J. Kruger, *Int. Mater. Rev.* **33**, 113 (1988).

24. Z. Szklarska-Smialowska, *Corrosion* **27**, 223 (1971).

25. H. H. Strehblow, *Werkstoffe Korrosion* **35**, 437 (1984).

26. W. H. J. Vernon and L. Whitby, *J. Inst. Met.* **42**, 181 (1929); **44**, 389 (1930).

27. W. H. J. Vernon, *J. Inst. Met.* **49**, 153 (1932).

28. J. C. Scully, *Corrosion Australasian* **7**, 4 (1982).

29. *Corrosion in Action*, 6th ed., The International Nickel Co., Inc., New York, 1977.

30. *Metals Handbook*, 9th ed., Vol. 13 (*Corrosion*), ASM International, Metals Park, Ohio, 1987.

31. *Metals Handbook*, 8th ed., Vol. 1, ASM International, Metals Park, Ohio, 1961, p. 1001.

32. H. S. Campbell, in W. H. Ailor, ed., *Handbook on Corrosion Testing and Evaluation*, John Wiley & Sons, Inc., New York, 1971, p. 5.

33. D. A. Jones, *Principles and Prevention of Corrosion*, Macmillan Publishing Co., New York, 1992, p. 4.

34. D. A. Jones, in Ref. 33, pp. 234–289.

35. E. H. Spuhler and C. L. Burton, *Avoiding Stress Corrosion Cracking in High Strength Aluminum Alloy Structures*, ALCOA Green Letter, Aluminum Company of America, Pittsburg, Pa., Apr. 1970.

36. *Corrosion Resistance of Nickel-Containing Alloys in Phosphoric Acid*, International Nickel Company; Corrosion Engineering Bulletin CEB-4, International Nickel Co., New York, 1961, p. 15.

37. N. Hackerman and E. S. Snively, in Ref. 6, p. 127.

38. E. McCafferty, in H. Leidheiser, Jr., ed., *Corrosion Control by Coatings*, Science Press, Princeton, N.J., 1979, p. 279.

39. N. E. Hamner, in Ref. 6, p. 245.

40. E. McCafferty, P. M. Natishan, and G. K. Hubler, *Nucl. Instr. Meth.* **B56**, 639 (1991).

Patrick J. Moran
United States Naval Academy

Paul M. Natishan
Naval Research Laboratory

CORUNDUM.

See ABRASIVES; ALUMINUM COMPOUNDS.

COSMETICS

Cosmetics are products created by the cosmetic industry and marketed directly to consumers. The cosmetic industry is dominated by manufacturers of finished products but also includes manufacturers who sell products to distributors as well as suppliers of raw and packaging materials. Cosmetics represent a large group of consumer products designed to improve the health, cleanliness, and physical appearance of the human exterior and to protect a body part against damage from the environment. Cosmetics are promoted to the public and are available without prescription.

The difference between a cosmetic and a drug is often confusing. In the United States the inclusion of a drug constituent, as defined by the Food and Drug Administration (FDA), in a cosmetic product may make the product a drug; whenever there is a claim for pharmacological activity of one of a product's constituents, the product is a drug. Some products are identified as quasi or over-the-counter (OTC) drugs according to each country's regulations. The composition, claim structure, and distribution of OTC products may be more tightly regulated than those of pharmacologically inactive cosmetics. The difference between an ordinary cosmetic and a quasi or OTC drug may not be readily apparent; it is based on statutory regulations. Certain types of products, such as hair-growth products and skin rejuvenators, are not cosmetics, and OTC claims for hair growth or skin rejuvenation are not allowed in the United States.

Cosmetics, regardless of form, can be grouped by product use into the following seven categories: (1) skin care and maintenance, including products that soften (emollients and lubricants), hydrate (moisturizers), tone (astringents), protect (sunscreens), etc, and repair (antichapping, antiwrinkling, antiacne agents); (2) cleansing, including soap, bath preparations, shampoos, and dentifrices (qv); (3) odor improvement by use of fragrance, deodorants, and antiperspirants; (4) hair removal, aided by shaving preparations, and depilatories; (5) hair care and maintenance, including waving, straightening, antidandruff, styling and setting, conditioning, and coloring products (see HAIR PREPARATIONS); (6) care and maintenance of mucous membranes by use of mouthwashes, intimate care products, and lip antichapping products; and (7) decorative cosmetics, used to beautify eyes, lips, skin, and nails.

History

Cosmetic preparations and usage are rooted in antiquity, when suspensions of natural pigments in lipids were evidently used to enhance appearance, and fragrant plant concoctions were widely traded. The use of cosmetics for adornment is recorded in biblical writings, and the use of soap (qv), probably a hydrolysate of animal lipids by wood ashes, was encouraged for cleanliness. The benefits of bathing were fully known to the ancients, who built elaborate bathhouses. Bathing became less popular in Western cultures during the Middle Ages but again became accepted during the eighteenth and nineteenth centuries.

The use of fragrant substances has been continuous, and the use of lipids or emollients for anointing is fully documented in historical writings. However, it is probably not justifiable to identify the recipes passed on from antiquity as cosmetics. The compositions based on folklore and mysticism were replaced by more scientifically acceptable products beginning about 1875. The first edition of a handbook of cosmetic chemistry published in 1920 included a foreword noting that scientific cosmetic chemistry did not exist prior to that publication (1). A few years later, texts on cosmetic chemistry and other formularies became available (2).

The Society of Cosmetic Chemists, with individual memberships, was founded in the United States after World War II, based on the belief that scientific expertise and exchange were the foundations for future expansion of the cosmetic industry. Prior to that time, knowledge of cosmetic formulation was jealously guarded. Related scientific societies emerged in other countries and have since joined to form the International Federation of Societies of Cosmetic Chemists.

Economic Aspects

Economic summaries of the cosmetic industry, commonly documented by sales volume, are sometimes based on unit sales, sometimes on manufacturers' sales in monetary units, and sometimes on consumer spending. Figures normally include contributions by private labeling operations but do not necessarily reflect the value of the industry service sector, which includes suppliers of raw materials, beauticians, testing laboratories, and other specialists. Moreover, product categories cannot be rigidly defined. For example, the differentiation between a deodorant (a cosmetic) and an antiperspirant (an OTC drug) is often obscured by its trade name.

A summary by broad categories is given in Table 1. A more detailed breakdown of U.S. cosmetic sales is provided in Table 2. The U.S. Commerce Department reports a modest (8%) increase in the value of cosmetic industry shipments between 1989 and 1991, despite a 13% decrease in total industry employment.

Table 1. U.S. Manufacturers' Sales of Cosmetics and Toiletries^a

Product line	Annual sales, \$ × 10 ⁶		Average annual growth, %
	1985	1990	
hair care	3,630	4,760	5.5
skin care	2,500	3,510	7.0
color cosmetics	2,540	3,380	5.9
fragrances	2,370	2,840	3.7
other toiletries	4,160	5,270	4.8
Total	15,200	19,760	5.4

^a Courtesy of Kline & Company, Fairfield, N.J.

Table 2. 1991 U.S. Cosmetic Sales^a

Product line	Sales, \$ × 10 ⁹
hair care	5.10
fragrances	3.90
makeup (color)	3.00
skin care	1.80
antiperspirant and deodorants	1.60
dentifrices	1.20
mouthwashes	0.58
shaving	0.58
hair coloring	0.57
sun care	0.40
nail care	0.37
<i>Total</i>	20.00

^a Based on U.S. Department of Commerce *Economic Industry Reports* and estimates by the staff of *Drug and Cosmetic Industry*.

Numerous cosmetic trade organizations exist. Foremost among them are the Cosmetic, Toiletry, and Fragrance Association (CTFA), formerly the Toilet Goods Association; the European Cosmetics Industries Federation (COLIPA); and the Japanese Cosmetic Industry Association (JCIA). These organizations provide member companies with regulatory and technical information and supply documentation on the industry's practices to governments and consumers. The cosmetic industry supports a number of trade journals. Comprehensive annual listings of companies, individuals, and products are available (3–5).

Regulation of the Cosmetic Industry

In the United States, the 1938 revision of the Federal Food and Drug Act regulates cosmetic products and identifies these materials as:

(1) articles intended to be rubbed, poured, sprinkled, or sprayed on, introduced into, or otherwise applied to the human body or any part thereof for cleansing, beautifying, promoting attractiveness, or altering the appearance, and (2) articles intended for use as a component of any such articles, except that such term shall not include soap.

This definition establishes the legal difference between a drug and a cosmetic. It is clearly the purpose of, or the claims for, the product, not necessarily its performance, that legally classifies it as a drug or a cosmetic in the United States. For example, a skin-care product intended to beautify by removing wrinkles may be viewed as a cosmetic because it alters the appearance and a drug because it affects a body structure.

The FDA is responsible for enforcing the 1939 act as well as the Fair Packaging and Labeling Act. In light of the difficulty of differentiating between cosmetics and drugs, the FDA has in recent years implemented its regulatory power by concluding that certain topically applied products should be identified as OTC drugs. As a group, these OTC drugs were originally considered cosmetics and remain among the products distributed by cosmetic companies. They include acne, antidandruff, antiperspirant, astringent, oral-care, skin-protectant, and sunscreen products.

The use or presence of poisonous or deleterious substances in cosmetics and drugs is prohibited. The presence of such materials makes the product "adulterated" or "misbranded" and in violation of good manufacturing practices (GMP), which are applicable to drugs and, with minor changes, to cosmetics (6).

In contrast to prescription drugs, OTC drugs and cosmetics are not subject to preclearance in the United States. However, the rules covering OTC drugs preclude introduction of untested drugs or new combinations. A "new chemical entity" that appears suitable for OTC drug use requires work-up via the new drug application (NDA) process. In contrast, the use of ingredients in cosmetics is essentially unrestricted and may include less well known substances.

Color Additives. The FDA has created a unique classification and strict limitations on color additives (see also COLORANTS FOR FOOD, DRUGS, COSMETICS, AND MEDICAL DEVICES). Certified color additives are synthetic organic dyes that are described in an approved color additive petition. Each manufactured lot of a certified dye must be analyzed and certified by the FDA prior to usage. Color lakes are pigments (qv) that consist of an insoluble metallic salt of a certified color additive deposited on an inert substrate. Lakes are subject to the color additive regulations of the FDA and must be certified by FDA prior to use. Noncertified color additives require an approved color additive petition, but individual batches need not be FDA certified prior to use.

Hair colorants, the fourth class of color additives, may be used only to color scalp hair and may not be used in the area of the eye. Use of these colorants is exempt, that is, coal-tar hair dyes may be sold with cautionary labeling, directions for preliminary (patch) testing, and restrictions against use in or near the eye. The FDA diligently enforces the rules governing color additives and limits the use of, or even delists colorants deemed unsafe. The list of substances specifically prohibited for use in cosmetics is short.

Under the Fair Packaging and Labeling Act, the FDA has instituted regulations for identifying components of cosmetics on product labels. To avoid confusion, the CTFA has established standardized names for about 6000 cosmetic ingredients (7). Rigid U.S. labeling requirements mandate that ingredients be listed in order of descending concentration. Similar regulations are expected for European cosmetics within the next few years (8).

European Regulations. Regulations for cosmetics differ from country to country but, in general, are similar to or patterned after U.S. regulation. Thus, the identification of a cosmetic in the European Community differs only marginally from that in the United States. A 1991 European Economic Community (EEC) directive defines a cosmetic as:

any substance or preparation intended for placing in contact with the various external parts of the human body (epidermis, hair system, nails, lips and external genital organs) or with the teeth and the mucous membranes of the oral cavity with a view to cleaning them, perfuming them, protecting them, keeping them in good condition, changing their appearance and/or correcting body odours.

EEC proposals also assert that cosmetic products must not damage human health when applied under normal or reasonably foreseeable conditions of use. Finally, the EEC directive states that the label of a cosmetic should include a list of ingredients in descending order of weight at the time of manufacture.

Although the EEC is still in the process of completing cosmetic regulation, the final directive is expected to require member states to ban marketing of cosmetics that contain prohibited ingredients, an amount of a substance in excess of that proscribed, coloring agents, preservatives, or uv filters not specifically allowed.

Japanese Regulation. Japanese regulations of cosmetics are similar to those already discussed. The safety and quality of cosmetic products are regulated under the Pharmaceutical Affairs Law with detailed requirements for approval and licensing of manufacturing and import, for labeling and advertising standards, and for reporting safety data to the Ministry of Health and Welfare.

Cosmetics in Japan are defined as externally used articles for cleaning, beautifying, promoting attractiveness, and altering the appearance of the

human body and for keeping the skin and hair healthy, provided that the action of the article on the human body is mild. Articles intended for use in diagnosis, treatment, or prevention of disease and those intended to affect the structure or any function of the body are identified as quasi drugs and are excluded. Japanese law identifies the following as quasi drugs: products for the prevention of foul breath or body odor; products for the prevention of prickly heat; products for the prevention of hair loss, promotion of hair growth, or removal of hair; hair dyes; agents for permanent waving of hair; and agents combining cosmetic effects with the purpose of preventing acne, chapping, itchy skin rashes, chilblains, or disinfection of the skin or mouth.

The Japanese regulations include both a list of substances that may not be used in cosmetics and a list of ingredients that may be used but which must conform to the specifications of the Japanese Standards of Cosmetic Ingredients. Some ingredients are allowed only in some types of cosmetic preparations. The use of certain hormones is controlled, and concentration limits exist for another group of ingredients. Japanese regulations differ significantly from U.S. regulations with regard to the formal recognition of an allowed, or positive, list.

Regulatory changes and discussions of the impact of regulations on the manufacture and import of cosmetic products are available in manuals published by the CTFA (9).

Product Requirements

Safety. Cosmetic products must meet acceptable standards of safety during use, must be produced under sanitary conditions, and must exhibit stability during storage, shipment, and use. Cosmetics are not lifesaving or life-prolonging drugs, and the requirements for innocuousness are absolute. In the United States the manufacturer bears the responsibility for not using injurious or questionable ingredients. The safety of each ingredient used in each finished cosmetic product must be adequately substantiated prior to marketing. In countries that have positive lists of ingredients that may be used in cosmetics, the burden for testing each finished cosmetic products is reduced. Positive listing assumes, without requiring evidence, that no adverse effects result from the use of a mixture of safe ingredients.

For many years the safety of cosmetic ingredients has been established using a variety of animal safety tests. More recently animal welfare organizations have urged that this type of safety testing be abandoned. Despite widespread use of cosmetics without professional supervision, the incidence of injury from cosmetic products is rare. In part, this is the result of extensive animal safety testing of components as well as of finished products. Such animal testing was considered mandatory from about 1945 to about 1985. Since the mid-1980s animal testing has been significantly reduced.

In vitro safety testing technology is becoming more common. Validation of these methods is based on comparisons with early animal safety data. Whether these *in vitro* tests can ensure the safety of all products that reach the consumer cannot be predicted (10). As of this writing, the principle that *in vitro* testing may be substituted for *in vivo* testing for complete safety substantiation has not been accepted by regulatory agencies. In the United States, the CTFA created the Cosmetic Ingredient Review (CIR) for the purpose of evaluating existing *in vitro* and *in vivo* data and reviewing the safety of the ingredients used in cosmetics. The review of ingredients is prioritized based on frequency of use, concentration used, the area of use, and consumer complaints. The CIR conclusions are available from the CTFA.

In addition to the CIR process the cosmetic industry has instituted a second, important, self-regulatory procedure: the voluntary reporting of adverse reactions, which is intended to provide data on the type and incidence of adverse reactions noted by consumers or by their medical advisors. This reporting procedure creates early awareness of problems handled outside hospital emergency facilities or centers for acute poisoning.

Safety testing of a finished cosmetic product should be sufficient to ensure that the product does not cause irritation when used in accordance with directions, neither elicits sensitization nor includes a sensitizer, and does not cause photoallergic responses. Some of the methods for determining animal or human responses to cosmetics are noted in Table 3.

Table 3. Cosmetic Safety Tests

Determination	Test vehicle	
	Animals	Humans
primary irritation	acute	chronic (21 days)
	albino rabbit with or without scarification ^a	occlusive; chamber scarification ^b Duhring soap chamber ^c
contact allergy	guinea pig maximization (with or without Freund's complement adjuvant) ^d	repeat patch test (open and closed); maximization test ^e
phototoxicity	hairless mouse and other species plus uvA irradiation ^f	human test ^{f,g}
photoallergy	guinea pig plus uvA + uvB irradiation ^h	photomaximization test ⁱ
eye irritation	Draize rabbit eye test ^j	eye instillation using diluted product
comedogenicity	rabbit ear ^k	in-use test
sting test		face ^{l,m}

^a Ref. 11.

^b Ref. 12.

^c Ref. 13.

^d Ref. 14.

^e Ref. 15.

^f Ref. 16.

^g Ref. 17.

^h Ref. 18.

ⁱ Ref. 19.

^j Ref. 20.

^k Ref. 21.

^l Ref. 22.

^m Ref. 23.

A particularly critical test for establishing the safety of cosmetics is the exaggerated-use test, in which panelists, often under medical supervision, use a product at frequencies that exceed the normally expected usage. Any adverse reactions, including subjective reports of burning or itching without clinical symptoms, suggest that the product should be examined further. This test also can be used to elicit comments concerning product acceptability.

Repeated usage of certain common cosmetic ingredients can elicit a response within the sebaceous gland apparatus that generates comedos. The cause of this phenomenon is not entirely clear, but an animal (rabbit ear) test purportedly measures the comedogenic potential of cosmetic ingredients or finished products (24). Controversy surrounds the identity of comedogenic substances and the concentration required to elicit the response. Thus use of

cosmetic ingredients that have been suspected of causing comedogenicity are generally avoided.

Production Facilities. The manufacture of acceptable cosmetic products requires not only safe ingredients but also facilities that maintain high standards of quality and cleanliness. Most countries have established regulations intended to assure that no substandard product or batch is distributed to consumers. Good Manufacturing Practices (GMP) represent workable standards that cover every aspect of drug manufacture, from building construction to distribution of finished products. GMPs in the United States that have been established for drug manufacture are commonly used in cosmetic production (6,25).

Contamination. Manufacturers of cosmetics must be careful to guard against chemical and microbial contamination. Chemical contamination, which may result from the presence of undesirable impurities in raw materials, is avoidable by adhering to rigid specifications for raw materials. Compendial specifications and publications by the CTFA and other professional societies form the basis of most intracompany raw material specifications. Moreover, all packaging components must meet not only physical and design specifications but also such chemical requirements as extractables and absence of dust and similar contaminants (see PACKAGING, COSMETICS AND PHARMACEUTICALS).

Chemical contamination arising from overheating or other decomposition reactions during processing or from improper storage of incoming supplies must also be avoided. For these reasons, adherence to documented production processes and periodic reassays of stored supplies are required. Additionally, final chemical or physical examinations of the finished and filled products are required to ascertain that no inadvertent chemical contamination has occurred during manufacture and that no undesirable ingredients are present.

An entirely different type of contamination arises from the presence of microbiota in a product. As in the case of chemical contamination, compendial requirements for microbiological purity exists. Pharmacopoeial standards vary from country to country, and manufacturers must use the specifications and kill times that meet local requirements. As of this writing, the criteria in the *British Pharmacopoeia* are more stringent than those established by the CTFA, which are stricter than those in the *United States Pharmacopoeia*. In order to meet commonly accepted standards of microbial purity, manufacturing facilities must be periodically cleaned and all products that can support microbial growth must contain an effective preservative (6).

Stability. An additional mandatory requirement for cosmetic products is chemical and physical stability. Interactions between ingredients that lead to new chemical entities or decomposition products are unacceptable. Stability testing becomes particularly critical if the product includes an active or drug constituent for which a specific performance claim is made. In the absence of an expiration date, a cosmetic product or an OTC drug should be stable for 60 months at ambient temperature. This temperature is a function of climatic zones. Therefore, controlled temperature storage, sometimes at controlled relative humidity, is universally recognized as ideal despite its attendant cost. In order to demonstrate long-term chemical stability on the basis of short- or intermediate-term studies, formulations are stored routinely at elevated temperatures, normally 37, 45, or 50°C. Changes are extrapolated to ambient temperatures using the Arrhenius equation for reaction rates.

Another type of chemical change is initiated by light, which may trigger autolytic, that is, free radical (Type I) or singlet oxygen (Type II) reactions. These changes are routinely classified as oxidation. Rancidity in cosmetics, especially those containing unsaturated lipids, is commonly prevented by use of antioxidants (qv).

Requirements for physical stability in cosmetics are not as rigid as those for chemical stability. As a rule, minor changes in viscosity or appearance are acceptable to users. More drastic changes, resulting from separation of an emulsion because of creaming or oiling, are not acceptable. Short-term physical, or viscosity, changes cannot be extrapolated to long-term performance. Changes observed during static viscosity tests have little predictive value for long-term viscosity or emulsion stability. Short-term dynamic viscosity tests also do not allow prediction of long-term viscosity changes, but these can sometimes be used to predict changes in the nature of emulsions. Zeta potential and particle size determination can provide predictive information on emulsion behavior.

Performance. Consumer acceptance is a criterion on which cosmetic marketers cannot compromise. Whereas the likes and dislikes of consumers are in a state of constant flux, some product features are critical. A deodorant that does not deodorize or a hair coloring that fades in sunlight is unacceptable. Performance is tested by *in vitro* techniques during formulation, but the ultimate test of a product's performance requires in-use experience with consumers and critical assessment by trained observers. Performance tests can sometimes be combined with in-use safety tests, and protocols for such programs have been developed.

Ingredients

Manufacturers of cosmetics employ a surprisingly large number of raw materials. Some of these ingredients are active constituents that have purported beneficial effects on the skin, hair, or nails, for example, acting as moisturizers or conditioners. These substances are generally used in limited quantities. Other ingredients are used to formulate or create the vehicle. These are bulk chemicals used in comparatively large amounts. The resulting combination of various substances affects the nature (viscosity, oiliness, etc) of the finished cosmetic. As a rule numerous combinations and permutations are tested to optimize textural characteristics and to match these to consumers' preferences. Finally, cosmetics may include substances added primarily to appeal to consumers. These ingredients need not contribute appreciably to product performance.

About 6000 different cosmetic ingredients have been identified (7). These can be divided into smaller groups according to chemical similarity or functionality. Table 4 represents a breakdown by functionality on the skin or in the product. The chemical identity of only one ingredient that performs the desired function is given. In most cases, other equally effective substances exist. The diversity of functions required in cosmetics is evident, and cosmetic ingredients may perform more than one function or belong to more than one chemical class. A typical example is sodium DL-2-pyrrolidinone-5-carboxylate (sodium PCA) [28874-51-3], NaC₅H₇NO₃. Chemically, this compound may be viewed as an amide, a heterocyclic compound, or an organic salt; functionally, it is a humectant and skin-conditioning agent.

Table 4. Cosmetic Functions and Representative Ingredients^a

Function	Ingredient ^b	Molecular formula	CAS Registry Number
<i>Biologically active agents</i>			
antiacne	salicylic acid	C ₇ H O ₃	[69-72-7]
anticaries	monosodium fluorophosphate	Na ₂ HPO ₃ F	[10163-15-2]
antidandruff	zinc pyrithione	C ₁₀ H ₈ N ₂ O ₂ S ₂ Zn	[13463-41-7]
antimicrobial	benzalkonium chloride		[8001-54-5]
antiperspirant	aluminum chlorohydrate	Al ₂ ClH ₅ O ₅	[12042-91-0]
biocides	triclosan	C ₁₂ H ₇ Cl ₃ O ₂	[3380-34-5]
sunscreen	octyl methoxycinnamate	C ₁₈ H ₂ O ₃	[5466-77-3]
skin protectant	dimethicone	(C ₂ H OSi) _n C ₄ H ₁₂ Si	[9006-65-9]
		(C ₂ H OSi) _n	[63148-62-9]
		C ₈ H ₈ O ₃	[9016-00-6]
external analgesic	methyl salicylate		[119-36-8]
<i>Nonbiologically active agents</i>			
abrasive			
skin	oatmeal		
teeth	dicalcium phosphate	Ca ₂ (HPO ₄) ₂	[7757-93-9]
antifoam	simethicone		[8050-81-5]
antioxidant	ascorbic acid	C H ₈ O	[50-81-7]

antistatic agent	dimethylditallow alkylammonium chlorides		[68783-78-8]
binder	hydroxypropylcellulose		[9004-64-2]
chelator	hydroxyethyl ethylenediamine triacetic acid (HEDTA)	C ₁₀ H ₁₈ N ₂ O ₇	[150-39-0]
colorant			
pigment	ultramarine	Na ₇ Al Si O ₂₄ S ₂	[1317-97-11]; [1345-00-2]; [12769-96-9]
dye	FD&C Red No. 4	C ₁₈ H ₁ N ₂ O ₇ S ₂ ·2Na	[4548-53-2]
emulsion stabilizer	xanthan gum		[11138-66-2]
film former	PVP	(C H ₉ NO) _x	[9003-39-8]
hair colorant	<i>p</i> -phenylenediamine	C H ₈ N ₂	[106-50-3]
hair conditioner	sodium lauroamphoacetate	Na ₂ C ₁₈ H ₃₅ N ₂ O ₃ ·HO	[14350-96-0]
humectant	glycerol	C ₃ H ₈ O ₃	[56-81-5]
deodorant			
mouth	zinc chloride	ZnCl ₂	[7646-85-7]
external	cetylpyridinium chloride	C ₂₁ H ₃₈ ClN	[123-03-5]
preservative	propylparaben	C ₁₀ H ₁₂ O ₃	[94-13-3]
emollient	octyl stearate	C ₂ H ₅₂ O ₂	[22047-49-0]
skin-conditioning agent			
general	pyrrolidinone carboxylic acid (PCA)	C ₅ H ₇ NO ₃	[98-79-3]
occlusive	petrolatum	C _n H _{2n+2}	[8009-03-8]
film forming	hyaluronic acid		[9004-61-9]
solvent	ethanol	C ₂ H O	[64-17-5]
cleansing agent	sodium lauryl sulfate	C ₁₂ H ₂₅ NaO ₄ S	[151-21-3]
emulsifying agent	polysorbate 65		[9005-71-4]
foam booster	cocamide DEA		[68140-00-1]
suspending agent	sodium lignosulfonate		[8061-51-6]
hydrotrope	sodium toluenesulfonate		[12068-03-0]
viscosity-controlling agent			
decrease	propylene glycol	C ₃ H ₈ O ₂	[57-55-6]
increase	hydroxypropylmethyl-cellulose		[9004-65-3]

^a Additional functions may be found in Ref. 26.
^b CTFA adopted names are used; this notation is used for cosmetic labeling.

Ingredients exhibiting certain functions are required in many types of cosmetic products. Antioxidants and preservatives are especially critical for product shelf life and quality during usage. Shelf life is defined herein as that period of time during which a product in an unopened package maintains its quality and performance and shows no physical or chemical instability. Antioxidants and preservatives do not contribute to physical stability but are included in cosmetic products to ensure oxidative stability and to control microbial contamination. Once a package has been opened, oxidative processes may cause the product to deteriorate, and microbial species may gain access to the product. These additives are expected to impart some protection even under these circumstances.

Antioxidants. Some antioxidants useful in cosmetics are listed in Table 5. The operant mechanisms are interference with radical propagation reactions, reaction with oxygen, or reduction of active oxygen species. Antioxidants are intended to protect the product but not the skin against oxidative damage resulting from ultraviolet radiation or singlet oxygen formation.

Table 5. Free-Radical-Inhibiting Antioxidants or Reductants Useful in Cosmetics^{a,b}

Antioxidant	CAS Registry Number	Molecular formula
ascorbic acid	[50-81-7]	C H ₈ O
ascorbyl palmitate	[137-66-6]	C ₂₂ H ₃₈ O ₇
butylated hydroxyanisole (BHA)	[25013-16-5]	C ₁₁ H ₁ O ₂
butylated hydroxytoluene (BHT)	[128-37-0]	C ₁₅ H ₂₄ O
<i>t</i> -butyl hydroquinone	[1948-33-0]	C ₁₀ H ₁₄ O ₂
cysteine	[52-90-4]	C ₃ H ₇ NO ₂ S
dilauryl thiodipropionate	[123-28-4]	C ₃₀ H ₅₈ O ₄ S
dodecyl gallate	[1166-52-5]	C ₁₉ H ₃₀ O ₅
ellagic acid	[476-66-4]	C ₁₄ H O ₈
erythorbic acid	[98-65-6]	C H ₈ O
kaempferol	[520-18-3]	C ₁₅ H ₁₀ O
nordihydroguaiaaretic acid	[500-38-9]	C ₁₈ H ₂₂ O ₄
propyl gallate	[121-79-9]	C ₁₀ H ₁₂ O ₅
quercetin	[117-39-5]	C ₁₅ H ₁₀ O ₇
sodium ascorbate	[134-03-2]	C H ₇ NaO
sodium sulfite	[7757-83-7]	Na ₂ SO ₃
thioglycolic acid	[68-11-1]	C ₂ H ₄ O ₂ S
tocopherol	[59-02-9]; [1406-18-4]	C ₂₈ H ₄₈ O ₂

^a Ref. 26 includes a more comprehensive listing.
^b Use levels are normally about 0.1% and rarely exceed 0.2%.

Preservatives. Several microorganisms can survive and propagate on unpreserved cosmetic products. Preservatives are routinely added to all preparations that can support microbial growth. The choice of a preservative for a given product is difficult. Anhydrous preparations and products containing high levels of ethanol or *i*-propanol may not require the addition of preservatives.

Contamination during manufacture is common, even when microbially clean ingredients are used. Water, which is almost ubiquitous in cosmetic products, is especially troublesome and must be free from contaminating microorganisms. All other ingredients should be screened for the presence of microbial species and batches of raw materials of dubious purity may have to be rejected. Cleanliness during manufacture, processing, and filling must be strictly maintained. Despite these precautions, microbial integrity of products may require the presence of one or more preservatives that are compatible with the product's ingredients. Products should not support the growth or viability of any microbial species that may have been accidentally introduced. Preservatives are also required to reduce contamination by consumers during normal use. Powerful preservative action to create self-sterilizing products is required. Whereas production of sterile cosmetics may be practicable, maintenance of sterility during use is problematical, because fingers and cosmetic applicators are not sterile.

Pharmacopoeias and CTFA publications provide guidelines for challenge test procedures and limits on microbial counts (25). The compendial requirements for kill of microorganisms vary significantly, and alternative test methods may be required (27). As a general rule, pathogenic organisms should be absent (28). Table 6 lists a number of antimicrobial preservatives used in cosmetic products. Experience has shown that some of the most commonly used preservatives are inactivated by a variety of surfactants. For example, the parabens (esters of *p*-hydroxybenzoic acid) are exceptionally sensitive to the presence of nonionic surfactants, presumably as a result of micellization of the antimicrobial by the surfactant. Over the years, preservation problems have resulted in the introduction into cosmetics of unusual substances that exhibit suitable antimicrobial spectra. However, some of these ingredients reportedly are irritants or sensitizers. Controversies in the scientific literature over the use of these substances are aggravated by regulatory acceptance or prohibition, which may differ from country to country. Table 6 includes preservatives that may be barred in certain countries.

Table 6. Antimicrobial Preservatives Useful in Cosmetics^{a, b}

Name	CAS Registry Number	Molecular formula
benzoic acid ^c	[65-85-0]	C ₇ H O ₂
benzyl alcohol	[100-51-6]	C ₇ H ₈ O
5-bromo-5-nitro-1,3-dioxane	[30007-47-7]	C ₄ H BrNO ₄
2-bromo-2-nitropropane-1,3-diol	[52-51-7]	C ₃ H BrNO ₄
butylparaben	[94-26-8]	C ₁₁ H ₁₄ O ₃
calcium propionate	[4075-81-4]	CaC H ₁₀ O ₄
chlorobutanol	[57-15-8]	C ₄ H ₇ Cl ₃ O
<i>m</i> -cresol	[108-39-4]	C ₇ H ₈ O
<i>o</i> -cresol	[95-48-7]	C ₇ H ₈ O
<i>p</i> -cresol	[106-44-5]	C ₇ H ₈ O
DEDM hydantoin	[26850-24-8]	C ₉ H ₁ N ₂ O ₄
dehydroacetic acid	[520-45-6]	C ₈ H ₈ O ₄
diazolidinyl urea	[278-92-2]	C ₁₁ H ₈ O ₂
dimethyl oxazolidine	[51200-87-4]	C ₅ H ₁₁ NO
DMDM hydantoin	[6440-58-0]	C ₇ H ₁₂ N ₂ O ₄
7-ethylbicyclooxazolidine	[7747-35-5]	C ₇ H ₁₃ NO ₂
ethylparaben	[120-47-8]	C ₉ H ₁₀ O ₃
formaldehyde	[50-00-0]	CH ₂ O
glutaral	[111-30-8]	C ₅ H ₈ O ₂
glyoxal	[107-22-2]	C ₂ H ₂ O ₂
imidazolidinyl urea	[39236-46-9]	C ₁₁ H ₁ N ₈ O ₈
iodopropynyl butylcarbamate	[55406-53-6]	C ₈ H ₁₂ INO ₂
isobutylparaben	[4247-02-3]	C ₁₁ H ₁₄ O ₃
isopropylparaben	[4191-73-5]	C ₁₀ H ₁₂ O ₃
MDM hydantoin	[116-25-6]	C H ₁₀ N ₂ O ₃
methylchloroisothiazolinone	[26172-55-4]	C ₄ H ₄ ClNOS
methyldibromoglutaronitrile	[35691-65-7]	C H Br ₂ N ₂
methylisothiazolinone	[2682-20-4]	C ₄ H ₅ NOS
methylparaben	[99-76-3]	C ₈ H ₈ O ₃
phenethyl alcohol	[200-456-2]	C ₈ H ₁₀ O
phenol	[108-95-2]	C H O
phenoxyethanol	[122-99-6]	C ₈ H ₁₀ O ₂
phenylmercuric acetate	[62-38-4]	HgC ₈ H ₈ O ₂
phenylmercuric benzoate	[94-43-9]	HgC ₁₃ H ₁₀ O ₂
phenylmercuric borate	[102-98-7]	HgC H ₇ BO ₃
<i>o</i> -phenylphenol	[90-43-7]	C ₁₂ H ₁₀ O
propylparaben	[94-13-3]	C ₁₀ H ₁₂ O ₃
Quaternium-14	[27479-28-3]	C ₂₃ H ₄₂ N ·Cl
Quaternium-15	[51229-78-8]	C ₉ H ₁ ClN ₄ ·Cl
sodium dehydroacetate	[4418-26-2]	NaC ₈ H ₇ O ₄
sodium phenolsulfonate	[1300-51-2]	NaC H ₅ O ₄ S
sodium phenoxide	[139-02-6]	NaC H ₅ O
sodium pyrrithione	[3811-73-2]	NaC ₅ H ₅ NOS
sorbic acid ^c	[110-44-1]	C H ₈ O ₂
thimerosal	[54-64-8]	NaHgC ₉ H ₉ O ₃ S
triclocarban	[101-20-2]	C ₁₃ H ₉ Cl ₃ N ₂ O
triclosan	[3380-34-5]	C ₁₂ H ₇ Cl ₃ O ₂
zinc pyrithione	[13463-41-7]	ZnC ₁₀ H ₈ N ₂ O ₂ S ₂

^a Ref. 26 includes a more comprehensive listing.
^b Use levels are product dependent but generally do not exceed 0.25%.

^c The acid salts are also used.

Local restrictions concerning the inclusion of preservatives and other constituents are dependent on the cosmetic product's method of use. Products that are allowed to remain on the skin are differentiated from those that are meant to be rinsed off. Components of products left on the skin can be expected to penetrate the viable epidermis and to be systematically absorbed. Products that are rinsed off shortly after skin contact, such as shampoos, can, if properly labeled, contain preservatives that might elicit adverse reactions if left on the skin. Typical examples of such preservatives are formaldehyde, formaldehyde releasers such as Quaternium 15 or MDM hydantoin, and the blend of methylchloroisothiazolinone and methylisothiazolinone.

Decorative eye cosmetic products have been reported to be subject to pathogenic microbial contamination. Regulatory agencies in several countries, therefore, permit the use of mercury-containing preservatives in eye makeups. The infections reported were to a large extent caused by contamination during use, and the introduction of self-sterilizing preparations seems warranted.

Lipids. Natural and synthetic lipids are used in almost all cosmetic products. Lipids serve as emollients or occlusive agents, lubricants, binders for creating compressed powders, adhesives to hold makeup in place, and hardeners in such products as lipsticks. In addition, lipids are used as gloss-imparting agents in hair-care products. The primary requirements for lipids in cosmetics are absence of excessive greasiness and ease of spreading on skin. Oily lipids, principal constituents of emulsions (creams and lotions), are well suited for inclusion in massage products, oils used to treat the skin (bath oils), ointments, suntan oils, and the like. Selection for a specific application is made on the basis of chemical inertness and physical properties. Petrolatum, mineral oils, polymeric silicones, polybutenes, and related substances are ingredients used for skin and hair conditioning. Conditioning is cosmetic jargon for describing a substance's beneficial effect on the substrate. For example, quaternary compounds are substantive to skin and hair proteins and thus can produce conditioning effects. Similarly, lipidic compounds without substantive functional groups, for example, tricaptin, condition skin merely by their presence on the surface. A selected listing of cosmetically useful lipids is provided in Table 7.

Table 7. Cosmetically Useful Lipids^a

Material	CAS Registry Number	Molecular formula
<i>Emollients</i>		
butyl oleate	[142-77-8]	C ₂₂ H ₄₂ O ₂
caprylic capric glycerides	[65381-09-1]	
cetyl lactate	[35274-05-6]	C ₁₉ H ₃₈ O ₃
dibutyl sebacate	[109-43-3]	C ₁₈ H ₃₄ O ₄
diisobutyl adipate	[141-04-8]	C ₁₄ H ₂₆ O ₄
ethyl linoleate	[544-35-4]	C ₂₀ H ₃₂ O ₂
glyceryl isostearate	[32057-14-0]	C ₂₁ H ₄₂ O ₄
hydrogenated palm kernel glycerides ^b		
isodecyl myristate	[17670-91-6]	C ₂₄ H ₄₈ O ₂
isopropyl stearate	[112-10-7]	C ₂₁ H ₄₂ O ₂
lauryl lactate	[6283-92-7]	C ₁₅ H ₃₀ O ₃
mineral oil	[8012-95-1]	C _{<i>n</i>} H _{<i>2n</i>}
myristyl myristate	[3234-85-3]	C ₂₄ H ₅₀ O ₂
oleyl oleate	[3687-45-4]	C ₃₇ H ₇₆ O ₂
PPG-10 cetyl ether	[9035-85-2]	(C ₃ H ₇ O) ₂ C ₁₁ H ₂₄ O
propylene glycol dicaprylate	[7384-97-6]	C ₁₅ H ₁₉ NOS HCl
squalene	[111-02-4]	C ₃₀ H ₅₀
wheat germ glycerides	[58990-07-8]	
<i>Occlusive agents</i>		
acetylated lanolin	[61788-48-5]	
butyl stearate	[123-95-5]	C ₂₂ H ₄₄ O ₂
caprylic capric triglyceride	[65381-09-1]	
dimethicone	[9006-65-9]	(C ₂ H ₅ OSi) _{<i>n</i>} C ₄ H ₁₂ Si
hydrogenated rice bran wax ^c		
lauryl stearate	[5303-25-3]	C ₃₀ H ₆₀ O ₂
paraffin	[8002-74-2]	C _{<i>n</i>} H _{<i>2n+2</i>}
pentarerythritol tetrastearate	[115-83-3]	C ₇₇ H ₁₄₈ O ₈
petrolatum	[8009-03-8]	C _{<i>n</i>} H _{<i>2n+2</i>}
propylene glycol dipelargonate	[225-350-9]	C ₂₁ H ₄₀ O ₄
stearyl erucate ^d		C ₄₀ H ₇₈ O ₂
trilinolein	[537-40-6]	C ₅₇ H ₉₈ O
<i>Natural lipids</i>		
apricot kernel oil	[72869-69-3]	
beeswax	[8006-40-4]	
carnauba	[8015-86-9]	
castor oil	[8001-79-4]	
coconut oil	[8001-31-8]	
japan wax	[8001-39-6]	
jojoba wax	[66625-78-3]	
lanolin	[8006-54-0]	
mink oil ^e		
olive oil	[8001-25-0]	
ozokerite	[8021-55-4]	
rice bran oil	[68553-81-1]; [84696-37-7]	
sesame oil	[8008-74-0]	
sunflower seed oil	[8001-21-6]	
vegetable oil	[68956-68-3]	
walnut oil	[8024-09-7]	

^a Ref. 26 includes a more comprehensive listing.

^b This is a hydrogenated mixture of mono-, di-, and triglycerides derived from palm kernel oil.

^c Prepared by partial hydrogenation of rice bran wax.
^d Erucic acid, *n*-octadecanol ester.
^e Oil obtained from subdermal fatty tissue of genus *Mustela*.

Solvents. Solvents can be added to cosmetics to help dissolve components used in cosmetic preparations. Water is the most common solvent and is the continuous phase in most suspensions and water oil (w/o) emulsions. Organic solvents are required in the preparation of colognes, hair fixatives, and nail lacquers. Selected solvents are used to remove soil, sebum, and makeup from skin. Solvents used in cosmetics include acetone, denatured alcohol, butoxyethanol (ethyleneglycol monobutylether), diethylene glycol, dimethyl isosorbide, ethyl acetate, heptane, isopropyl alcohol, mineral spirits (boiling range 110–155°C), polyethylene glycol (mol wt from 200 up to 15,000), propylene glycol, toluene, and tricaprin (glyceryl tri-*n*-decanoate). A comprehensive listing may be found in Reference 26. The selection of solvents for use in cosmetics is a complex task because of odor as well as topical and inhalation toxicities.

Surfactants. Substances commonly classified as surfactants (qv) or surface active agents are required in a wide variety of cosmetics. These are often categorized on the basis of ionic character but are grouped in Table 8, which includes at least one member from each of the various chemical types of surfactants, on the basis of utility in cosmetics. Prolonged contact with anionic surfactants can cause some swelling of the skin. Although this is a temporary phenomenon, skin in this swollen condition allows permeation of externally applied substances. Nonionic surfactants as a group are generally believed to be mild even under exaggerated conditions. The more hydrophobic nonionics, those that are water dispersible (not water-soluble), can enhance transdermal passage. Amphoteric surfactants as a group exhibit a favorable safety profile. Finally, cationic surfactants are commonly rated as more irritating than the anionics, but the evidence for generalized conclusions is insufficient.

Table 8. Cosmetic Surfactants^a

Material ^b	CAS Registry Number	Molecular formula
<i>Cleansing agents</i>		
ammonium laureth sulfate ^{c,d}	[32612-48-9]	(C ₂ H ₄ O) _{<i>n</i>} C ₁₂ H ₂ O ₄ S·H ₃ N
cetalkonium chloride	[122-18-9]	C ₂₅ H ₄ N·Cl
DEA myristate	[53404-39-0]	C ₁₄ H ₂₈ O ₂ ·C ₄ H ₁₁ NO ₂
decyl polyglucose ^{d,e}		
dioctyl sodium sulfosuccinate ^{c,d}	[577-11-7]	C ₂₀ H ₃₈ O ₇ S·Na
disodium cocoamphodiacetate ^d	[68650-39-5]	
disodium laurimino dipropionate ^d	[3655-00-3]	C ₁₈ H ₃₅ NO ₄ ·2Na
lauryl betaine ^{c,d}	[683-10-3]	C ₁ H ₃₃ NO ₂
lauryl pyrrolidone ^d	[2687-96-9]	C ₁ H ₃₁ NO
nonoxynol-12	[9016-45-9]	(C ₂ H ₄ O) _{<i>n</i>} C ₁₅ H ₂₄ O
myristamine oxide ^{c,d}	[3332-27-2]	C ₁ H ₃₅ NO
PEG-50 stearate	[9004-99-3]	(C ₂ H ₄ O) _{<i>n</i>} C ₁₈ H ₃ O ₂
potassium dodecylbenzenesulfonate ^d	[27177-77-1]	KC ₁₈ H ₃₀ O ₃ S
potassium oleate	[143-18-0]	KC ₁₈ H ₃₄ O ₂
sodium cocoyl glutamate ^d	[68187-32-6]	
sodium C ₁₄₋₁ olefin sulfonate ^d	[68439-57-6]	
sodium laureth phosphate ^{c,d}	[42612-52-2]	
sodium lauryl sulfate ^{c,d}	[151-21-3]	NaC ₁₂ H ₂ O ₄ S
sodium methyl oleoyl taurate ^c	[137-20-2]	NaC ₂₁ H ₄₁ NO ₄ S
sodium nonoxynol-25 sulfate	[9014-90-8]	(C ₂ H ₄ O) _{<i>n</i>} C ₁₅ H ₂₄ O ₄ S·Na
sodium oleoyl isethionate ^d	[142-15-4]	NaC ₂₀ H ₃₈ O ₅ S
sodium stearate ^c	[822-16-2]	NaC ₁₈ H ₃ O ₂
TEA-abietoyl hydrolyzed collagen ^d	[68918-77-4]	
TEA-lauryl sulfate ^d	[139-96-8]	C ₁₂ H ₂ O ₄ S·C H ₁₅ NO ₃
TEA-oleoyl sarcosinate ^c	[17736-08-2]	C ₂₁ H ₃₉ NO ₃ ·C H ₁₅ NO ₃
<i>Emulsifying agents</i>		
ceteareth-10	[68439-49-6]	
cetrimonium bromide	[57-09-0]	C ₁₉ H ₄₂ N·Br
laneth-5	[3055-95-6]	C ₂₂ H ₄ O
lecithin	[8002-43-5]	
nonoxynol-9	[14409-72-4]	C ₃₃ H ₀ O ₁₀
PEG-20 dilaurate	[9005-02-1]	(C ₂ H ₄ O) _{<i>n</i>} C ₂₄ H ₄ O ₃
PEG-8 oleate	[9004-96-0]	(C ₂ H ₄ O) _{<i>n</i>} C ₁₈ H ₃₄ O ₂
poloxamer 407	[9003-11-6]	(C ₃ H O·C ₂ H ₄ O) _{<i>x</i>}
polyglyceryl-8 oleate	[9007-48-1]	
polysorbate 60	[9005-67-8]	
sorbitan sequioleate	[8007-43-0]	
sucrose stearate	[25168-73-4]	C ₃₀ H ₅ O ₁₂
<i>Foam boosters</i>		
cocamine oxide	[61788-90-7]	
lauramide DEA	[120-40-1]	C ₁ H ₃₃ NO ₃
myristamide MIPA	[10525-14-1]	C ₁₇ H ₃₅ NO ₂
myristaminopropionic acid	[14960-08-8]	C ₁₇ H ₃₅ NO ₂
<i>Hydrotropes</i>		
ammonium xylenesulfonate	[26447-10-9]	C ₈ H ₁₀ O ₃ S·H ₃ N
potassium toluenesulfonate	[16106-44-8]	C ₇ H ₈ O ₃ S·K
sodium methyl naphthalene sulfonate	[26264-58-4]	C ₁₁ H ₁₀ O ₃ S·Na
<i>Solubilizing agents</i>		

cetareth-40	[68439-49-6]	
oleth-44	[9004-98-2]	(C ₂ H ₄ O) _{<i>n</i>} C ₁₈ H ₃ O
PEG-40 stearate	[9004-99-3]	(C ₂ H ₄ O) ₂ C ₁₈ H ₃ O ₂
	<i>Suspending agents</i>	
behentrimonium chloride	[17301-53-0]	C ₂₅ H ₅₄ N ·Cl
benzethonium chloride	[121-54-0]	C ₂₇ H ₄₂ NO ₂ ·Cl
sodium lignosulfonate	[8061-51-6]	
sodium polystyrene sulfonate	[9003-59-2]	(C ₈ H ₈ O ₃ S ·Na) _{<i>x</i>}

^a Ref. 26 includes a comprehensive listing.

^b CFTA names are used.

^c Belongs to a chemical class especially useful in facial and body washes.

^d Belongs to a chemical class especially useful in shampoos.

^e Decylether of a glucose oligomer.

Colorants. Color (qv) is used in cosmetic products for several reasons: the addition of color to a product makes it more attractive and enhances consumer acceptance; tinting helps hide discoloration resulting from use of a particular ingredient or from age; and finally, decorative cosmetics owe their existence to color.

Organic Colorants. The importance of coal-tar colorants cannot be overemphasized. The cosmetic industry, in cooperation with the FDA, has spent a great deal of time and money in efforts to establish the safety of these dyes (see COLORANTS FOR FOOD, DRUGS, COSMETICS, AND MEDICAL DEVICES). Contamination, especially by heavy metals, and other impurities arising from the synthesis of permitted dyes are strictly controlled. Despite this effort, the number of usable organic dyes and of pigments derived from them has been drastically curtailed by regulatory action.

In addition to the U.S. certified coal-tar colorants, some noncertified naturally occurring plant and animal colorants, such as alkanet, annatto [1393-63-1], carotene [36-88-4], C₄₀H₅₆, chlorophyll [1406-65-1], cochineal [1260-17-9], saffron [138-55-6], and henna [83-72-7], can be used in cosmetics. In the United States, however, natural food colors, such as beet extract or powder, turmeric, and saffron, are not allowed as cosmetic colorants.

The terms FD&C, D&C, and External D&C (Ext. D&C), which are part of the name of colorants, reflect the FDA's colorant certification. FD&C dyes may be used for foods, drugs, and cosmetics; D&C dyes are allowed in drugs and cosmetics; and Ext. D&C dyes are permitted only in topical products. Straight colorants include both the organic dyes and corresponding lakes, made by extending the colorant on a substrate such as aluminum hydroxide or barium sulfate. The pure dye content of these lakes varies from 2 to 80% ; the organic dyes contain over 80% pure dye. Colorants certified for cosmetic use may not contain more than 0.002% of lead, not more than 0.0002% of arsenic, and not more than 0.003% of heavy metals other than lead and arsenic.

Inorganic Colorants. In addition to various white pigments, other inorganic colorants such as those listed in Table 9 are used in a number of cosmetic products. These usually exhibit excellent lightfastness and are completely insoluble in solvents and water.

Table 9. Inorganic Pigments Useful in Makeups

Material	Molecular formula	Color
titanium dioxide	TiO ₂	white
zinc oxide	ZnO	white
talc	steatite	whitish
barium sulfate	BaSO ₄	white
mica		glossy, colorless
titanium dioxide–ferric oxide coated mica	^a	glossy, nacreous, multicolored
		multicolored
guanine ^b	C ₅ H ₅ N ₅ O	nacreous
bismuth oxychloride	BiOCl	white, nacreous
iron oxides	85% Fe ₂ O ₃	yellow to orange
umber	Fe ₂ O ₃ Fe ₃ O ₄	brown
sienna	Fe ₂ O ₃ (ignited)	red
	Fe ₃ O ₄	black
chrome hydroxide green	Cr ₂ O(OH) ₄	bluish green
chrome oxide greens	Cr ₂ O ₃	green
ferric ammonium ferrocyanide	Fe(NH ₄)[Fe(CN)]	blue
ferric ferrocyanide	Fe ₄ [Fe(CN)] ₃	blue
manganese violet	Mn(NH ₄)P ₂ O ₇	violet
ultramarines ^c		blue, violet, red, pink, green

^a Material is a mixture.

^b Guanine [73-40-5], an organic dye, is also known as CI 75170 [73-40-5].

^c Materials are fusion mixtures.

Naturally occurring colored minerals that contain oxides of iron are known by such names as ochre [1309-37-1], umber [12713-03-0], sienna [1309-37-1], etc. These show greater variation in color and tinting power than the synthetic equivalents, and the nature and amount of impurities in the national products is also variable. Most of the pigments identified in Table 9 are, therefore, manufactured synthetically. They are primarily used in skin-makeup products and in eye-area colorants.

Nacreous Pigments. For many years nacreous pigments were limited to guanine (from fish scales) and bismuth oxychloride. Mica, gold, copper, and silver, in flake form, can also provide some interesting glossy effects in products and on the face. Guanine is relatively costly, and bismuth oxychloride darkens on exposure to light and is difficult to suspend because of its high specific gravity. An entirely new set of colored, iridescent, inorganic pigments, which may be described as mixtures of mica and titanium dioxide (sometimes with iron oxides), has been created by coating mica flakes with titanium dioxide. The wavelengths of light reflected from these compositions can produce a complete range of colored interference patterns. The particle size of the mica must be controlled and may not exceed 150 μm, at least in the United States. Additional color effects can be created by sandwiching the mica, TiO₂, and Fe₂O₃.

Specialized Cosmetic Technologies

Several specialized technologies have been perfected for cosmetic products. Among these, emulsification, stick technology, and powder blending are prominent.

Emulsification.

Emulsification is essential for the development of all types of skin- and hair-care preparations and a variety of makeup products. Emulsions (qv) are fine dispersions of one liquid or semisolid in a second liquid (the continuous phase) with which the first substance is not miscible. Generally, one of the phases is water and the other phase is an oily substance: oil-in-water emulsions are identified as o/w; water-in-oil emulsions as w/o. When oil and water are mixed by shaking or stirring in the absence of a surface-active agent, the two phases separate rapidly to minimize the interfacial energy. Maintenance of the dispersion of small droplets of the internal phase, a requirement for emulsification, is practical only by including at least one surface-active emulsifier in the oil-and-water blend.

The addition of emulsifiers (see Table 8) lowers the energy of the large interfacial area created by forming a huge number of small droplets from a single large drop. In practical emulsification technology, this thermodynamic emulsion stabilization is augmented by two other features. One is the formation of a rigid interfacial film on the surface of the droplets of the internal phase (29). This film, sometimes exhibiting the optical characteristics of a liquid crystal, acts as a mechanical barrier to the coalescence of the droplets of the internal phase. Finally, the droplets may be stabilized by the formation of an electric double layer, which favors the electrical repulsion between charged particles. The latter requires the presence of an electrolyte or an ionized emulsifier.

The coalescence of internal phase droplets can be further decreased by raising the viscosity of the external continuous phase through addition of gums or synthetic polymers, for example, cellulosic gums such as hydroxypropyl methyl-cellulose [9004-65-3], fermentation gums such as xanthan gum [11138-66-2], or cross-linked carboxyvinyl polymers such as carbomer [39007-16-3]. The increased viscosity also counteracts changes in the emulsion resulting from differences in the specific gravity of the two phases as mandated by Stokes' law. An advance in cosmetic emulsification technology has resulted from the development of cross-linked carboxyvinyl polymers, in which some of the carboxylic acid residues are esterified with various fatty alcohols. These polymers possess the ability to act as primary emulsifiers and thicken the system when some of the remaining carboxylic groups are neutralized with alkali (see CARBOXYLIC ACIDS).

The selection of emulsifiers, auxiliary emulsifiers, gums, and other components is complicated and largely empirical. Despite the lack of a rigid theoretical basis, the hydrophile/lipophile balance (HLB) is the most useful approach for the selection of nonionic emulsifiers (30). The inclusion of ionic emulsifiers was not contemplated in the original formulation of the HLB system. The HLB system also does not account for the effect of low HLB viscosity increasing ingredients, such as cetyl alcohol or glyceryl monostearate. The precise selection of the desired blend from commercial nonionics for emulsification is often frustrating (31). Methods for selecting suitable blends of emulsifiers and stabilizers (32,33), for preparing emulsions (34,35), and for studying stability (36) have been published. The technical literature also includes publications dealing with the theory of emulsification and the structure of emulsions and of microemulsions (37,38).

Conventional cosmetic emulsions (macroemulsions) normally contain about 70% or more of the external phase, which may be a mixture of components. The internal phase is routinely introduced into the external phase at an elevated temperature with vigorous agitation. The emulsifiers are distributed according to their solubility between the two phases. The level of emulsifiers (rarely more than about 10%) is kept low since excessive amounts may destabilize emulsions or form a clear solubilizate. Auxiliary emulsifiers and other components are included in the phases in which they are soluble.

The term multiple emulsion describes a w/o emulsion in an o/w emulsion. For example, when a w/o emulsion is added to water, no dispersion is expected unless the aqueous phase is fortified with a suitable emulsifier. The resulting dispersion may then be a blend of a w/o and an o/w emulsion, or it may be a multiple emulsion of the w/o/w type. In this latter case, the initial w/o emulsion becomes the internal phase of the final product. Generally, these preparations are not very stable unless they are produced under rigidly controlled conditions (32,39,40).

Microemulsions or solubilized or transparent systems are very important in the marketing of cosmetic products to enhance consumer appeal (32,41). As a rule, large quantities of hydrophilic surfactants are required to effect solubilization. Alternatively, a combination of a solvent and a surfactant can provide a practical solution. In modern clear mouthwash preparations, for example, the flavoring oils are solubilized in part by the solvent (alcohol) and in part by the surfactants. The nature of solubilized systems is not clear. Under normal circumstances, microemulsions are stable and form spontaneously. Formation of a microemulsion requires little or no agitation. Microemulsions may become cloudy on heating or cooling, but clarity at intermediate temperatures is restored automatically.

Formation of liposomal vesicles under controlled conditions of emulsification of lipids with phospholipids has achieved prominence in the development of drugs and cosmetics (42). Such vesicles are formed not only by phospholipids but also by certain nonionic emulsifying agents. Formation is further enhanced by use of specialized agitation equipment known as microfluidizers. The almost spontaneous formation of liposomal vesicles arises from the self-assembly concepts of surfactant molecules (43). Vesicles of this type are unusual sustained-release disperse systems that have been widely promoted in the drug and cosmetic industries.

Stick Technology. Cosmetic sticks can be divided into three categories: sticks molded in the container; sticks molded separately and then encased; and sticks formed by compression.

Container Molding. Antiperspirant, deodorant, sunscreen, and antiacne sticks are container molded. The amount of dispersed ingredients makes them brittle and difficult to handle mechanically.

The required solids are suspended in a wax–emollient blend at about 60 to 80°C and milled. The liquid suspension is then cooled to about 55°C, and the more volatile ingredients are added. The mass is placed into containers, which are commonly provided with a threaded shaft for raising or lowering the product.

Antiperspirant sticks based on this molding technique have become more popular since volatile low mol wt cyclomethicones [69430-24-6] have been used successfully as the lipids and fatty alcohols as the waxes. This type of product delivers the active antiperspirant to the site as a clinging powder without excessive oiliness.

Deodorant and cologne sticks are formed by allowing sodium stearate to gel in a suitable organic solvent, usually ethanol or propylene glycol. The soap and the solvent are heated under reflux until the soap is dissolved. The solution is cooled to about 60°C; fragrance, color, and the like are added; and the mass is placed into suitable containers.

Stick Molding. Various types of lipsticks and eye-shadow sticks are stick molded. A wax-containing lipid mass is milled with the pigment at elevated (about 75°C) temperatures until it is uniform. The lipid mass, at a temperature about 10°C above its melting point, is then poured into metallic molds. Deaeration is essential to prevent unsightly depressions on the molded sticks. To avoid sudden congealing, the molds are customarily heated to a temperature above room temperature before filling; after filling, they are chilled to temperatures well below room temperature. After unmolding, the sticks may be inserted into various types of containers (swivel, metal, or plastic). In order to formulate acceptable molded sticks, slowly developing surface anomalies or defects must be avoided. Foremost are the excrescence of solid fatty substances (also called bloom) and the exudation of liquid substances (also called sweating). Both of these defects are attributed to polymorphic transformation. The selection of the proper blend of lipids to create an acceptable makeup stick is complex. Some of the lipids used in such products and their primary characteristics are listed in Table 10. Other types of sticks, for example, eyebrow pencils and lip liners, are molded similarly but may be inserted into wooden pencil stock, trimmed, and appropriately finished.

Table 10. Lipids Used in Cosmetic Molded Sticks^a

Lipid	CAS Registry Number	Characteristics
<i>Oils</i>		
castor oil	[8001-79-4]	high gloss; high viscosity

mineral oil	[8012-95-1]	high gloss
Esters/Alcohols		
butyl stearate	[123-95-5]	rapid wetting of pigments; controls sweating at elevated temperatures
guerbet alcohols		
decyl tetradecanol	[58670-89-6]	satiny gloss
octyl dodecanol	[5333-42-6]	satiny gloss
isocetyl alcohol	[36311-34-9]	similar to castor oil
isopropyl myristate	[110-27-0]	same as butyl stearate
isopropyl palmitate	[142-91-6]	same as butyl stearate
oleyl alcohol	[143-28-2]	high gloss; turns rancid
Fats		
cocoa butter	[8002-31-1]	tendency to bloom
glyceryl monostearate	[31566-31-1]	high melting
hydrogenated cocoglycerides ^b		variable properties
hydrogenated vegetable oil	[68334-28-1]	greasy
lanolin	[8006-54-0]	wets pigments
Waxes		
beeswax	[8006-40-4]	low gloss
candelilla	[8006-44-8]	low melting
carnauba	[8015-86-9]	very hard
ozokerite	[8021-55-4]	thermally stable
Dye solvents		
ethoxydiglycol acetate	[112-15-2]	bitter
polyethylene glycols	[25322-68-3]	
propylene glycol	[57-55-6]	

^a Ref. 26 includes comprehensive listings.

^b These are blends of mono-, di-, and triglycerides of hydrogenated coconut oil.

The criteria for a good cosmetic makeup stick are manifold: the sticks must pay off, that is, deliver the desired amount when used at or near room temperature; sticks must withstand exposure to moderately elevated temperatures; they should not break during normal use; stick components must not elicit unpleasant, for example, oily or warming, sensations after application; pigments must be uniformly distributed and must not react photochemically with the remaining stick components to cause rancidity or photoirritation; and finally, the film produced on the body site must be resistant to rubbing off or transferring to eating utensils or clothing.

Compressed-Powder Sticks. Compression of a blend of solids using a suitable binder or by extruding a water containing magma results in compressed-powder sticks.

Powder Blending. Cosmetic powders serve two primary functions. One group, commonly called body powders or talcs, is applied to the skin to provide lubricity and to absorb excessive moisture. The second group, commonly referred to as face powders, exists in both loose and compressed forms and is used to impart some color to the skin and to dull excessive oiliness.

Most powders, including medicated powders, depend on talc to provide lubricity and a matte finish on the skin. Talc is generally blended with other constituents, such as those listed in Table 11. The plate-like nature of mined talc makes this hydrous magnesium silicate (steatite) an important skin-care constituent. Loose body and makeup powders utilize additional bulk ingredients. The products can also include antimicrobial agents, dyes, and pigments. The selection of a fragrance must be made with great care; some bulk ingredients are alkaline, and the perfume oil on the surface of particles is subject to oxidation, especially if pigmented ingredients are included.

Table 11. Powder Ingredients Used for Cosmetics

Ingredient	Chemical identity	CAS Registry Number	Comment
chalk		[13397-25-6]	opaque, alkaline
kaolin (clay)	aluminum silicate	[1332-58-7]	
	attapulgite	[1337-76-4]	
	fuller's earth	[8031-18-3]	opaque, low gloss
	hectorite	[12173-47-6]	
	montmorillonite	[1318-93-0]	
magnesium carbonate		[546-93-0]	absorbent
metallic soaps	magnesium stearate	[557-704-0]	hydrophobic, lubricant
	zinc ricinoleate	[13040-19-2]	
silica	fumed	[7631-86-9]	absorbent
	xerogel	[112945-52-5]	
starch	corn starch	[9005-25-6]	hygroscopic
	rice starch	[9005-25-8]	
talc	hydrated magnesium silicate	[14807-96-6]	opaque lubricant
	(steatite)		
titanium dioxide		[13463-67-7]	opaque, white
zinc oxide		[1314-13-2]	opaque, adherent
zirconium silicate		[10101-52-7]	opaque, white

The basic manufacturing process involves thorough blending of the components, especially the pigments, and comminution with the aid of a variety of mills to reduce the particle size. Loose powders are filled without additional processing.

If compression is required to provide a stick or pan-type of product, the bulk components must be held together with a binder. Common binders are various lipids, polymers, polysaccharides, and waxes. Some binder compositions include water, which is removed by drying the compact. The amount of binder must be carefully controlled to yield a solid, nonfragile compact that is soft enough to pay off. Excessive amounts of or improperly compounded binders glaze during use because of transfer of skin lipids to the compact.

When the bulk containing the binder is uniform, it is compressed on pneumatic, hydraulic, or ram-type presses. Compression can be carried out in presses provided with suitably designed cavities or in metallic pans. The pans are filled with the powder mass, and a plunger with a cross-sectional shape similar to that of the pan is used to compress the tablet. The resulting tablets are commonly used with powder puffs or cosmetic brushes.

Skin Preparation Products

Products for use on the skin are designed to improve skin quality, to maintain (or restore) skin's youthful appearance, and to aid in alleviating the symptoms of minor diseases of the skin. Many of these products are subject to different regulations in different countries. Skin products are generally formulated for a specific consumer purpose.

Skin-Care Products. Preparations are generally classified by body part and purpose. For example:

Product	Purpose
baby preparations	
oils and lotions	cleansing, soothing
diaper-rash products	prevention, cure
powders	drying
foot preparations	
antifungals	anti-infective
emollients	soothing, crack-prevention
powders	drying
facial preparations	
lotions and creams	smoothing, protecting, rejuvenating
body preparations	
lotions and creams	smoothing, protecting
oils	smoothing, protecting
powders	drying
hand preparations	
lotions and creams	smoothing, protecting
gels	antichapping

The smoothing or emollient properties of creams and lotions are critical for making these emulsions the preferred vehicles for facial skin moisturizers, skin protectants, and rejuvenating products. On the body, emollients provide smoothness and tend to reduce the sensation of tightness commonly associated with dryness and loss of lipids from the skin. Although a wide variety of plant and animal extracts have been claimed to impart skin benefits, valid scientific evidence for efficacy has been provided only rarely.

Emulsion components enter the stratum corneum and other epidermal layers at different rates. Most of the water evaporates, and a residue of emulsifiers, lipids, and other nonvolatile constituents remains on the skin. Some of these materials and other product ingredients may permeate the skin; others remain on the surface. If the blend of nonvolatiles materially reduces the evaporative loss of water from the skin, known as the transepidermal water loss (TEWL), the film is identified as occlusive. Application of a layer of petrolatum to normal skin can reduce the TEWL, which is normally about 4–8 g (m²h), by as much as 50 to 75° for several hours. The evaporated water is to a large extent trapped under the occlusive layer hydrating or moisturizing the dead cells of the stratum corneum. The flexibility of isolated stratum corneum is dependent on the presence of water: dry stratum corneum is brittle and difficult to stretch or bend. Thus, any increase in the water content of skin is believed to improve the skin quality.

The ability to moisturize the stratum corneum has also been claimed for the presence of certain hydrophilic polymers, for example, guar hydroxypropyl trimonium chloride [65497-29-2], on the skin. By far the most popular way to moisturize skin is with humectants, some of which are listed in Table 12. It is claimed that humectants attract water from the environment and thereby provide moisture to the skin.

Table 12. Skin Conditioners and Moisturizers ^a

Material	CAS Registry Number	Molecular formula
glycerol	[56-81-5]	C ₃ H ₈ O ₃
2-pyrrolidinone-5-carboxylic acid (PCA)	[98-79-3]	C ₅ H ₇ NO ₃
sodium lactate	[72-17-3]	C ₃ H ₅ NaO ₃
urea	[57-13-6]	CH ₄ N ₂ O
cholesterol	[57-88-5]	C ₂₇ H ₄ O
hydrolyzed glycosaminoglycans ^b		
hydrolyzed soy protein	[68607-88-5]	
linoleic acid	[60-33-3]	C ₁₈ H ₃₂ O ₂
tocopheryl acetate	[7695-91-2]	C ₃₁ H ₅₂ O ₃
witch hazel distillate ^c		
sodium hyaluronate ^d		
myristyl betaine	[2601-33-4]	C ₁₈ H ₃₇ NO ₂

^a Ref. 26 includes a more comprehensive listing.

^b Mixed polysaccharides from animal connective tissue.

^c Nonalcoholic steam distillate of parts of *Hamamelis virginiana*.

^d Sodium salt of hyaluronic acid [9004-61-9].

Studies of the interactions between water and the lipid constituents of the stratum corneum suggest that the supply of water per se is not responsible for skin quality and condition. Water vapor from lower layers provides a constant supply of moisture to the epidermis. Instead, the ability of the skin to retain the moisture is critical, and this ability depends on the lipid lamellar bilayers that occupy the spaces between the cells of the stratum corneum (44–46).

In the United States, products claimed to reverse or alleviate the stigmata of facial skin aging are considered drugs. Claims for improvement of fine wrinkling, mottled hyperpigmentation, and roughness associated with photodamage, on the part of products containing all *trans*-retinoic acid [302-79-4], have received some favorable comments from regulatory advisory panels. Other approaches for anti-aging products are based on desquamation by α-hydroxyacids, for example, lactic acid [50-21-5]. Finally, a number of substances, such as hyaluronic acid [9004-61-9] and collagen [9007-34-5], have been claimed to improve the appearance of wrinkled skin (see also ANTIAGING AGENTS).

The amounts and types of lipids used in skin-care products control their application properties. Methods for assessing these characteristics using expert panelists have been described (47).

The ability of skin-care products to supply moisture to the skin remains in question. In the United States, however, the OTC panel has sanctioned the use of skin-protectant ingredients such as glycerin, which may play roles in the skin's water ecology. Products for the care of body skin are similar to preparations formulated for the care of facial skin. Products for overall body care should leave a dry, satinlike finish even though relatively high levels of

unctuous lipids are used. Facial night creams may leave the skin somewhat oily, whereas facial day creams must provide a dry finish. Hand-care products are designed to reduce chapping and cracking, especially prevalent during cold, dry, winter weather. Hand-care products are commonly fortified with various humectants, and products for the elbows and feet may include abrasives. Bath powders impart lubricity to body skin, absorb moisture, and provide some fragrance. These are formulated without pigments to preclude the staining of clothing. Skin-care formulations have been published for skin protectants (48), face creams (49), hand creams (50), and body creams (51).

Antiacne Preparations. Antiacne products are designed to alleviate the unsightly appearance and underlying cause of juvenile acne. Generally, acne is a mild disease of the follicular duct in which sebaceous secretion is not readily allowed to pass to the surface of the skin because of a hyperkeratotic restriction in the duct. The retained sebum may undergo chemical changes or be altered by microbial species, with consequent inflammatory responses. In the past, cosmetic preparations were designed to remove sebum from the skin surface with solvents or cleansers and work against microorganisms with antibacterials. In addition, acne was cosmetically treated with abrasives in the hope that scrubbing would relieve the ductal blockage.

As of 1991 in the United States, OTC antiacne preparations may contain only a few active drugs, for example, sulfur [7704-34-9], resorcinol acetate [102-29-4], resorcinol [108-46-3], salicylic acid [69-72-7], and some combinations (52). OTC anti-acne constituents may be included in a variety of conventional cosmetic preparations, which then become OTC drugs. These include lotions, creams, solutions, facial makeups, facial cleansers (including abrasive cleansers), and astringents. Products must contain the specified drugs at the designated concentrations. Compositions of antiacne products have been published (53).

Sunscreens. Radiation that reaches the earth's surface from the sun is limited to wavelengths above about 285 nm because shorter wavelengths are absorbed by ozone (qv). Investigations of the impact of ultraviolet (uv) light on human skin have identified the range from 285 to about 320 nm as uvB and that between 330 nm and visible light as uvA. Both uvA and uvB have the potential to burn skin, resulting in an acute sunburn that is painful and can have damaging long-term effects, such as wrinkling, actinic keratoses, or carcinomas. The flux of uvA and uvB that reaches human skin depends on altitude and latitude. Equatorial regions receive maximum flux. Clouds, dust, and reflected uv light from the ground or water also affect flux and may provide some wavelength specificity to the radiation. The flux required to produce a barely observable erythema, the so-called minimal erythema dose (MED), depends on the energy of the uv radiation. Thus, much higher fluxes of uvA than of uvB are required for the production of one MED. The estimation of the MED (a rapid skin response) is not a quantitative measure of long-term skin damage, especially because the lower energy uvA penetrates much deeper into the skin than does the shorter wavelength uvB.

The use of uv light absorbing substances is accepted worldwide as a means of protecting skin and body against damage and trauma from uv radiation. These colorless organic substances are raised to higher energy levels upon absorption of uv light, but little is known about mechanisms for the disposal of this energy. These substances can be classified by the wavelengths at which absorbance is maximal. Absorption throughout the incident uv range (285 to about 400 nm) affords the best protection against erythema.

It is also possible to deflect uv radiation by physically blocking the radiation using an opaque makeup product. A low particle size titanium dioxide can reflect uv light without the undesirable whitening effect on the skin that often results from products containing, for example, zinc oxide or regular grades of titanium dioxide.

In vitro absorption-spectrophotometry techniques are available to assess a sunscreen's efficacy, but the preferred methods are *in vivo* procedures in which a small body site is irradiated with the desired wavelengths for different periods in the presence or absence of a uv protectant. Procedures vary from country to country to determine the incremental timing of the exposure that ultimately allows quantification via sun protective factor (SPF). In the United States, sunscreen preparations are considered OTC drug products, and the SPF must be specified (54). Even in countries that do not identify these products as drugs, SPF labeling has become customary.

The SPF is defined as the ratio of the time required to produce a perceptible erythema on a site protected by a specified dose of the uv protectant product to the time required for minimal erythema development in the unprotected skin. An SPF of 8 indicates that the product allows a subject to expose the protected skin 8 times as long as the unprotected skin to produce the minimum erythema response. The measurement can be quite subjective unless skin color and the history of reactions to sun exposure of the test subjects are taken into account. The MED range for Caucasians at 300 nm averages 34 mJ cm². The range is 14–80 mJ cm². Perspiration or the use of artificial irradiation devices can create additional problems.

Because perspiration and bathing are commonly associated with sun exposure, the need to determine the SPF after bathing or long after application to the body site is important. In use, the quantity of screen applied and its uniform distribution over the exposed area control the achieved SPF. Methods for assessing the water-resistant or waterproof qualities of sunscreen products have been established by the FDA.

A list of uv absorbing substances found useful in protective sunscreen products is provided in Table 13. Some information on the levels permitted in products in both the United States and the EEC is included. Descriptions and specifications of sunscreens have been published (55).

Table 13. Cosmetic Uv Absorbers

Ingredient	CAS Registry Number	Quantity approved, %	
		U.S. ^a	EEC ^b
<i>uvA Absorbers</i>			
benzophenone-8	[131-53-3]	3	
menthyl anthranilate	[134-09-6]	3.5–5	
benzophenone-4	[406545-6]	5–10	5
benzophenone-3	[113-57-7]	2–6	10
<i>uvB Absorbers</i>			
<i>p</i> -aminobenzoic acid (PABA)	[150-13-0]	5–15	5
pentyl dimethyl PABA	[14779-78-3]	1–5	5
cinoxate	[104-28-9]	1–3	5
DEA <i>p</i> -methoxycinnamate		8–10	8
digalloyl trioleate	[17048-39-4]	2–5	4
ethyl dihydroxypropyl PABA	[5882-17-0]	1–5	5
octocrylene	[6187-30-4]	7–10	
octyl methoxycinnamate	[5460-77-3]	2–7.5	10
octyl salicylate	[118-60-5]	3–5	5
glyceryl PABA	[136-44-7]	2–3	5
homosalate	[118-56-9]	4–15	10
lawsone (0.25%)	[83-72-7]		
plus dihydroxyacetone (33%)	[96-26-4]		
octyl dimethyl PABA	[21245-02-3]	1.4–8	8
2-phenylbenzimidazole-5-sulfonic acid	[27503-81-7]	1–4	8
TEA salicylate	[2174-16-7]	5–12	2
sulfomethyl benzylidene bornanone	[90457-82-2]		10
urocanic acid (and esters)	[104-98-3]		2
<i>Physical barriers</i>			
red petrolatum		30–100	

titanium dioxide	[13463-67-7]	2–25
^a Ref. 55.		
^b Tentative.		

In principle, emulsified sunscreen products are similar to emollient skin-care products in which some of the emollient lipids are replaced by uv absorbers. The formulation of an effective sunscreen product generally requires combination of a uvB and a uvA absorber if an SPF above about 12 is desired. Two or more of the sunscreens listed in Table 13 normally constitute about one-half of the nonvolatiles found in sunscreen lotions. The other half consists of an emollient (solvent) and emulsifying and bodying agents. If water-resistant qualities are desired, polymeric film formers, for example, acrylates–octylacrylamide copolymers [9002-93-1], or water-repellent lipids, for example, dimethicone [9006-65-9], are included.

More recently anhydrous sunscreens have become popular. Products of this type are based on blends of emollient lipids and acceptable uv absorbers. Formulations of sunscreen products have been published (56).

Facial Makeup. This classification applies to all products intended to impart a satinlike tinted finish to facial skin and includes liquid makeups, tinted loose or compressed powders, rouges, and blushers.

In modern liquid makeups and rouges, the required pigments are extended and ground in a blend of suitable cosmetic lipids. This magma is then emulsified, commonly as o w, in a water base. Soaps, monostearates, conditioning lipids, and viscosity-increasing clays are primary components. Nonionic emulsifiers can replace part or all of the soap. Pigment levels in these emulsions are about 15% but can range as high as 60% if the stabilizing clays are included. The viscosity of these types of preparations varies from fairly thin fluids to thixotropic viscous lotions to firm creams. These products must not dry too rapidly to permit spreading on the skin and feathering of the edges for proper shading.

Tinted dry powders form the second type of facial makeup. Commonly, the blended solids are compressed into compacts. The finished products, sold as compressed powders, rouges, or blushers, are applied to the face with the aid of powder puffs, brushes, or similar devices. Facial makeup compositions have been published for rouge (57), powder (58), and makeup (59).

Skin Coloring and Bleaching Preparations. Products designed to simulate a tan, to lighten skin color in general, or to decolorize small hyperpigmented areas such as age spots either impart to or remove color from the skin. Skin stains are intended to create the appearance of tanned skin without exposure to the sun. The most widely used ingredient is dihydroxyacetone [96-26-4] (2–5% at pH 4 to 6) which reacts with protein amino groups in the stratum corneum to produce yellowish brown Maillard products. Lawsone [83-72-7] and juglone [481-39-0] are known to stain skin directly. Stimulation of melanin formation is another approach to artificial tanning. Commercialization, which is limited, depends primarily on topical application of products containing tyrosine or a tyrosine recursor.

The number of cosmetically acceptable bleaching ingredients is very small, and products for this purpose are considered drugs in the United States. The most popular ingredient is hydroquinone [123-31-9] at 1–5%; the addition of uv light absorbers and antioxidants reportedly helps to reduce color recurrence. Effective bleaching requires repeated localized applications of a cream or ointment type of preparation. Formulations for skin lighteners (60) and skin darkeners (61) have been published.

Astringents

Astringents are designed to dry the skin, denature skin proteins, and tighten or reduce the size of pore openings on the skin surface. These products can have antimicrobial effects and are frequently buffered to lower the pH of skin. They are perfumed, hydro-alcoholic solutions of weak acids, such as tannic acid or potassium alum, and various plant extracts, such as birch leaf extract. The alcohol is not only a suitable solvent but also helps remove excess sebum and soil from the skin. After-shave lotions generally function as astringents.

In the United States, some astringents, depending on product claims, are considered OTC drugs (62). Only three ingredients, aluminum acetate [139-12-8], aluminum sulfate [10043-01-3], and hamamelis [84696-19-5], are considered safe and effective.

Antiperspirants and Deodorants. There are many forms of antiperspirants and deodorants: liquids, powders, creams, and sticks. Deodorants do not interfere with the delivery of eccrine or apocrine secretions to the skin surface but control odor by reodorization or antibacterial action. Deodorant products, regardless of form, are antimicrobial fragrance products. An important antimicrobial or cosmetic biocide used in many products is triclosan [3380-34-5]. Other active agents include zinc phenolsulfonate [127-82-2], *p*-chloro-*m*-xlenol [88-04-0], and cetrimonium bromide [57-09-0]. There have been claims that ion-exchange polymers and complexing agents provide protection against unpleasant body odors. In addition, delayed-release, that is, liposomal or encapsulated, substances of diverse activity have been employed.

The mechanism of antiperspirant action has not been fully established but probably is associated with blockage of ducts leading to the surface by protein denaturation by aluminum salts. The FDA has mandated that an antiperspirant product must reduce perspiration by at least 20% and has provided some guidelines for testing finished products. Some antiperspirant chemicals are listed in Table 14 (63).

Table 14. Antiperspirant Ingredients

Name	CAS Registry Number	Molecular formula
<i>Aluminum chlorohydrates</i> ^a		
aluminum chlorohydrate	[1327-41-9]; [12042-91-0]	Al ₂ (OH) ₅ Cl · <i>n</i> H ₂ O
aluminum sesquichlorohydrate	[11097-68-0]	Al ₂ (OH) _{4.5} Cl _{1.5} · <i>n</i> H ₂ O
aluminum dichlorohydrate	[1327-41-9]; [12042-91-0]	Al ₂ (OH) ₄ Cl ₂ · <i>n</i> H ₂ O
<i>Aluminum zirconium chlorohydrates</i> ^{a,b}		
aluminum zirconium octachlorohydrate		Al ₈ Zr(OH) ₂₀ Cl ₈ · <i>n</i> H ₂ O
aluminum zirconium pentachlorohydrate		Al ₈ Zr(OH) ₂₃ Cl ₅ · <i>n</i> H ₂ O
aluminum zirconium tetrachlorohydrate	[57158-29-9]	Al ₄ Zr(OH) ₁₂ Cl ₄ · <i>n</i> H ₂ O
aluminum zirconium trichlorohydrate		Al ₄ Zr(OH) ₁₃ Cl ₃ · <i>n</i> H ₂ O
<i>Aluminum salts</i>		
aluminum chloride	[7446-70-0]	AlCl ₃
buffered aluminum sulfate	[10043-01-3]; [18917-91-4]	Al ₂ (SO ₄) ₃ and Al(CH ₃ CHOHCOO) ₃

^a Partially dehydrated derivatives complexed with polyethylene glycol or propylene glycol exist. In the United States, derivatives in which some of the water of hydration has been replaced by glycine are particularly popular. Aluminum zirconium tetrachlorohydrax gly and related derivatives can be used in OTC antiperspirant products.

^b The Al :Zr ratio is variable.

Clear solutions of antiperspirants have been on the market for about 100 years. Cream and lotion types are o w emulsions commonly formulated using nonionic emulsifiers to avoid aluminum salt formation, especially by carboxylic acids. Cream antiperspirants are generally distributed in jars, whereas lotions are dispensed from roll-on types of containers.

Antiperspirant aerosols (qv) can be wet or dry. In the wet type, the antiperspirant chemical is dissolved in a suitable solvent, such as water–ethanol,

combined with emollients and so on, and dispensed after pressurization using an acceptable propellant. Dry aerosols may be based on a finely milled antiperspirant component suspended with emollients and suspending agents in a volatile liquid that is lost after dispensing.

Most antiperspirant sticks are molded. Sticks dominate in the U.S. market, whereas lotion and cream antiperspirants are preferred in Europe. Stick antiperspirant products may include suspending agents, coupling agents to wet the antiperspirant chemical (about 20–25° o), and emollients. The blend is prepared at about 65°C and poured at about 55°C. Antiperspirant (64) and deodorant (65) compositions have been published.

Cleansing Preparations

Cleansing preparations are products, based on surfactants or abrasives, that are designed to remove unwanted soil and debris from skin, hair, and the oral cavity. Soaps (qv) are the best-known cleansers but are not considered cosmetic products unless they are formulated with agents that prevent skin damage or contain antimicrobial agents. Soaps are the least costly and most popular skin cleansers available. Use as hair cleansers is limited, however, by the tendency to form insoluble alkaline earth soaps, which leave a dulling film on hair, and the taste of soaps generally precludes use in oral-care preparations. Soaps dominate the skin cleanser market, although they have been shown in closed patch skin tests to cause some irritation owing to their alkalinity. Modern skin cleansers (liquids and bars) for sensitive skin employ various synthetic detergents (see Table 8).

Skin Cleansers. Their mildness, foaming qualities, water solubility, and tolerance of slightly acid conditions (pH 5–6) make many of the surfactants listed in Table 8 attractive for use in formulating facial and body cleansers. Irritant qualities of preparations based on one or more of these surfactants can be further modified by the addition of lipids or agents that lower the defatting (drying) tendencies of the finished product.

The solubility characteristics of sodium acyl isethionates allow them to be used in synthetic detergent (syndet) bars. Complex blends of an isethionate and various soaps, free fatty acids, and small amounts of other surfactants reportedly are essentially nonirritant skin cleansers (66). As a rule, the more deterusive surfactants, for example alkyl sulfates, α-olefin sulfonates, and alkylaryl sulfonates, are used in limited amounts in skin cleansers. Most skin cleansers are compounded to leave an emollient residue on the skin after rinsing with water. Free fatty acids, alkyl betaines, and some compatible cationic or quaternary compounds have been found to be especially useful. A mildly acidic environment on the skin helps control the growth of resident microbial species. Detergent-based skin cleansers can be formulated with abrasives to remove scaly or hard-to-remove materials from the skin.

Foaming bath and shower preparations are based on blends of surfactants and various conditioning agents, many of which are derived from plants, and may contain a relatively high percentage of fragrance. Surfactants are those identified in Table 8 as being beneficial in facial and body washes. Foam boosters are used to create the billowing foam desired in many of these products; sequestering agents are added to prevent the formation of alkali metal salts, especially when soap is used.

Cream-type skin cleansers have been used for many years, particularly on the face. The classical cold cream consists of mineral oil (50–60° o), beeswax (~15%), borax (~1%), and water (30–40° o). Neutralization of the cerotic acid in beeswax by borax yields the emulsifier to form the cream. When applied to the face, the mineral oil acts as a solvent for sebum, soil, and makeup. The remains are tissueed off, leaving the skin clean with a lubricating, oily finish. This very simple composition can be modified with additional emulsifiers, thickening agents, and cosmetic additives. There are numerous cosmetic wipe-off cleansers, although the popularity of these oil-based products has declined. When still more emulsifiers, especially water-soluble nonionics, are included, the original wipe-off products are converted into the more popular rinse-off types. The principal cleansing agents in these products are the oily components; in contrast with the detergent-based cleansers, the surfactant is included only to aid in removal of the product during rinsing, not for detergency.

Hair Cleansers. Except for a few specialty preparations, hair cleansers, or shampoos, are based on aqueous surfactants. The detergents of choice in shampoos are listed in Table 8. The most popular surfactants in shampoos are alkyl sulfates and alkylether sulfates, commonly used at about 10–15° o active. These ingredients by themselves do not provide the dense, copious foam desired by consumers, and additives are required, especially for use on oily hair or scalps. The foam boosters usually found in finished shampoos are fatty acid alkanolamides, fatty alcohols, and amine ocides. In the so-called superamide foam boosters the ratio of alkanolamine to fatty acid is 1:1. In addition to about 1–2° o of an amide, most commonly the diethanolamide of lauric acid [120-40-1] or of coconut fatty acids [61791-31-9], almost all shampoos contain a hair-conditioning agent. The detergent removes lipids from the hair's surface, leaving hair with limited gloss and often difficult to comb. Hair that has been defatted by detergents has a tendency to retain a static electric charge. The resulting fly-away hair can be avoided by making the hair more conductive by means of moisture or electrolytes or by lubricating the hair with a lipid or conditioning agent. The preferred conditioning agents are materials substantive to hair, that is, not readily rinsed off by water.

Excessive degreasing by shampoos can be overcome by treatment with an after-shampoo (cream) rinse or a hair dressing. It is considered desirable to control the effects of excessive surface degreasing at the time of shampooing. A wide variety of hair-conditioning additives has been recommended and tested. Only a few have gained wide acceptance, for example:

Name	CAS Registry Number
dialkyl (C ₁₂ –C ₁₈) dimethylammonium chloride, hydrolyzed collagen	eg, [53401-74-9] [9015-54-7]
polymeric quaternary derivatives	eg, [26590-05-6] or [53568-66-4]
potassium cocoyl hydrolyzed collagen	[68920-65-0]
sodium cocoamphoacetate	eg, [68334-21-4]
sodium lauroyl glutamate	[22923-31-7]
stearamidopropyl betaine	[68920-65-0]

Modern shampoos containing one or more of these conditioners have been designated as two-in-one products. They provide good cleaning and leave hair conditioned without the need for a second treatment. Eye stinging and irritation caused by shampoos can be reduced by including nonionic surfactants (with 10–45 polyoxyethylene groups) or by adding an amphocarboxylate.

Dandruff, a benign scaling skin disease of the scalp, is commonly viewed as a hair problem. The etiology and therapy of dandruff are similar to those of seborrheic dermatitis (67). Antidandruff shampoos are formulated using antimicrobial or desquamating agents to reduce the lipophilic yeasts (qv) widely believed to be the cause of scalp flaking. In the United States, shampoos for which antidandruff claims are made are OTC drugs. The choice of active agents is limited to coal tar [8007-45-2], zinc pyrithione [13463-41-7], salicylic acid [69-72-7], selenium sulfide [7488-56-4], and sulfur [7704-34-9], which can be added to shampoos or other scalp preparations. Some of the fungicidal azoles and piroctone olamine (1-hydroxy-4-methyl-6-(2,4,4-tri-methylpentyl)-2-(1H) pyridinone, ethanolamine salt [68890-66-4] are active against these causative organisms but are not recognized in United States OTC regulations.

Oral Cleansing Products. Toothpastes and mouthwashes are considered cosmetic oral cleansers as long as claims about them are restricted to cleaning or deodorization. Because deodorization may depend on reduction of microbiota in the mouth, several antimicrobial agents, either quaternaries, such as benzethonium chloride [121-54-0], or phenolics, such as triclosan [3380-35-5], are permitted. Products that include anticaries or antigingivitis agents or claim to provide such treatment are considered drugs.

Mouthwashes are hydro-alcoholic preparations in which flavorants, essential oils (see OILS, ESSENTIAL), and other agents are combined to provide long-term breath deodorization. Palatability can be improved by including a polyhydric alcohol such as glycerin or sorbitol (see ALCOHOLS, POLYHYDRIC). Occasionally, anionic and nonionic surfactants are used to help solubilize flavorants and to help remove debris and bacteria from the mouth.

Dentifrices (qv), or toothpastes, depend on abrasives to clean and polish teeth. The principal ingredients in toothpastes or powders are 20–50° o polishing agents, such as calcium carbonate, di- or tricalcium phosphate, insoluble sodium metaphosphate, silica, and alumina; 0.5–1.0° o detergents, for example, soap or anionic surface-active agents; 0.3–10° o binders (gums); 20–60° o humectants, such as glycerol, propylene glycol, and sorbitol; sweeteners

(saccharin, sorbitol); preservatives, such as benzoic acid or *p*-hydroxybenzoates; flavors, for example, essential oils; and water. The most widely used surfactant is sodium lauryl sulfate, which is available with high purity. It produces the desired foam during brushing, acts as a cleansing agent, and has some bactericidal activity.

Transparent dentifrices can be prepared from certain xerogel silicas through use of high levels of polyhydric alcohols. Clarity depends on matching the refractive indexes of the silica and the liquid base. Compositions for liquid facial cleansers (68), shampoos (69), conditioning shampoos (70), dandruff shampoos (71), surfactant bars (72), toothpastes (73), and mouthwashes (74) have been published.

Shaving Products

Cosmetic shaving products are preparations for use before, during, or after shaving.

Preshaves. Preshave products are used primarily for dry (electric) shaving. Solid preshaves are usually compressed-powder sticks based on lubricating solids, such as talc or zinc or glyceryl stearate. Liquid preshaves are intended to remove perspiration residues and tighten and lubricate the skin. The alcohol content is relatively high (50–80° o) to accelerate drying. The remaining ingredients may be polymeric lubricants, such as 1–2° o polyvinylpyrrolidinone (PVP) [9003-39-8], emollients, such as 1–5° o diisopropyl adipate [6938-94-9], and up to about 5° o propylene glycol.

Shaving Creams. Despite the replacement of soap by synthetic detergents, products for wet shaving continue to be based on soap. Shaving creams and soaps are available as solids, that is, bars; creams, generally in tubes; or aerosols. Solids are essentially pure soaps applied to the face as foams with a brush. Shaving creams may be nonlathering (emulsion) and rarely consist entirely of lubricating lipids.

The principal ingredients of shaving creams and aerosols are liquid soaps, usually a blend of potassium, amine, and sodium salts of fatty acids, formulated to create a foam with the desired consistency and rinsing qualities. The soap blend may include synthetic surfactants, skin-conditioning agents, and other components. Modern nonaerosol shaving creams may contain 20–30° o soap (potassium or triethanolamine (TEA)), up to about 10° o glycerine, emollients, and foam stabilizers. Aerosol shaving creams are dilute forms of the cream types and are dispensed from the container with the aid of hydrocarbon propellants (up to about 10° o). Aerosol shaving creams may also include some emulsifiers to ensure uniform emulsification of the propellant during the short shaking period before dispensation.

The objectives of shaving creams include protecting the face from cuts by cushioning the razor. Beard-softening qualities are attributable almost exclusively to hair hydration, which also depends on pH. Lubrication must be provided, primarily between the blade and the hair fibers through which it passes. Blade technology based on polymeric coatings reduces the need for this type of lubrication by shaving creams.

After-Shaves. After-shave preparations serve the same function as and are formulated similarly to skin astringents. After-shave balms are hydro-alcoholic or alcohol-free emulsions that supply soothing ingredients, for example, witch hazel, and emollients, for example, decyl oleate [3687-46-5], to the skin. Menthol, which provides a cooling sensation, is a common constituent of after-shaves.

Compositions for preshaves (75), shaving creams (76), and after-shaves (77) have been published.

Nail-Care Products

Over the years the cosmetic industry has created a wide variety of products for nail care. Some of these, such as cuticle removers and nail hardeners, are functional; others, such as nail lacquers, lacquer removers, and nail elongators, are decorative.

Functional Nail-Care Products. Cuticle removers are solutions of dilute alkalis that facilitate removal, or at least softening, of the cuticle. Formulations containing as much as 5° o potassium hydroxide have been reported. Such preparations may contain about 10° o glycerine to reduce drying, and thickeners, such as clays, to reduce runoff. Lipids and other conditioners are included to reduce damage to tissues other than the cuticle.

Nail hardeners have been based on various protein cross-linking agents. Only formaldehyde is widely used commercially. Contact with skin and inhalation must be avoided to preclude sensitization and other adverse reactions. The popularity of products of this type is decreasing because the polymers used in nail elongators can be used to coat nails to increase the mechanical strength.

Decorative Nail-Care Products. Nail lacquers, or nail polishes, consist of resin, plasticizer, pigments, and solvents. The most commonly used resin is nitrocellulose, prepared by esterification of celluloses with nitric acid, with a degree of substitution between 1.8 and 2.3 nitrate groupings per anhydroglucose unit. Ethylacetate, butylacetate, and toluene are typical solvents. Toluenesulfonamide–formaldehyde resin [2503 5-71-6] and similar polymers, for example, the terpolymer of 2,2,4-trimethyl- 1,3-pentanediol, isophthalic acid, and trimellitic anhydride, are the resins of choice as secondary film formers for optimal nail adhesion. Other resins, such as alkyds, acrylates, and polyamides, can also serve as secondary film formers.

Camphor, dibutyl phthalate [84-74-2], and other lipidic solvents are common plasticizers. Nail lacquers require the presence of a suspending agent because pigments have a tendency to settle. Most tinted lacquers contain a suitable flocculating agent, such as stearalkonium hectorite, a reaction product of hectorite [12173-46-6] and stearalkonium chloride [122-19-0].

The blend of pigments used to create a particular shade must conform to regulations covering pigments and dyes in cosmetics. Regulations vary among countries and undergo frequent changes. The selected pigments may not stain the nails, and any organic dye or pigment that might exhibit solubility in the mixture of lacquer solvents is avoided. Typical organic dyes include monoazo dyes, such as D&C Red No. 6 Barium Lake and D&C Red No. 34 Calcium Lake, and pyrazole dyes, such as FD&C Yellow No. 5 Aluminum Lake (see AZO DYES). Inorganic pigments, such as iron oxides and titanium dioxide, can be incorporated. Colored pigments do not usually exhibit the opacity and reflective brilliance demanded of modern nail enamels. Bismuth oxychloride, mica, and guanine were extensively used in the past to provide nacreous reflections, whereas titanium dioxide provided opacity. These substances have been replaced by the highly reflective and nacreous synthetic mica–titanium dioxide pigments. Generally, nacreous pigments are supplied as suspensions in nitrocellulose-containing solvent blends, whereas other pigments are mixed with nitrocellulose and plasticizers and processed through a roller mill. Nitrocellulose processing must be done with extreme caution in an explosion-proof environment. The level of pigment in nail enamels generally does not exceed about 5° o. The solvent level is about 60–70° o, however, and the dried lacquer may contain as much as 10–15° o pigment.

Nail lacquer removers are simply acetone or blends of solvents similar to those used in nail lacquers. It is commonly accepted that solvents have a drying effect on nails, and nail lacquer removers are often fortified with various lipids such as castor oil [8991-79-4] or cetyl palmitate [540-10-3].

Nail elongators are products intended to lengthen nails. These have become extremely popular. In earlier compositions, polymerization was conducted by mixing monomers, oligomers, and catalysts on the nail (78). More recently, nail elongation is achieved by adhering a piece of non-woven nylon fabric (referred to as nail wrap) to the nail with a colorless lacquer. This process may be repeated until the desired nail thickness has been reached. After shaping, the artificial nail is further decorated. Compositions for nail lacquers (79) and nail hardeners and conditioners (80) have been published.

Hair Products

Cosmetics for hair care fall into several categories: cleansers or shampoos, conditioners, fixatives, coloring products, waving and straightening products, and hair removers (see HAIR PREPARATIONS).

Hair Conditioners. Hair conditioners are designed to repair chemical and environmental damage, replace natural lipids removed by shampooing, and facilitate managing and styling hair. The classical hair-conditioning products were based on lipids, which were deposited on hair either directly, with oils or pomades, or from emulsions. Liquid and semisolid brillantines are formulated from mineral oil or vegetable or animal fats thickened with waxes (ozokerite), fatty alcohols (cetyl alcohol), or polymers (for example, polyethylene), and are normally dispensed from jars or tubes. Emulsion products are commonly based on an oil phase that consists of mineral oil, lanolin, and synthetic or vegetable-derived lipids. The emulsifiers vary widely and may include anionics (soap), nonionics (alkyl polyoxyethylene ethers), or cationics (eg, PEG-2 stearmonium chloride [606087-87-8]). These o w type emulsions may be thickened with various gums and may contain plant extracts, antimicrobial agents (quaternaries), uv screens, and hair-fixative polymers such as PVP.

Microemulsions, temporary emulsions, that is, two-layer hair dressings, and clear solutions of nonvolatile lubricants are on the market. Hair tonics, usually hydro-alcoholic, achieve similar effects by including lipid substances or synthetic emollients, such as the mono butyl ethers of polypropylene oxides [9003-13-8] (10–50 mol). The primary benefits of these lipid-based products are lubrication and improvements in hair gloss and hair-holding (dressing)

qualities. Hair holding and manageability result from the tendency of the lipid components to make the fibers adhere to each other laterally, not from the coating of individual hairs.

An entirely different, and in the 1990s more popular, type of hair conditioning is achieved by treating hair with substantive quaternary compounds or quaternary polymers. Quaternaries are sorbed by hair, retained despite rinsing with water, and removed only by shampooing. Quaternaries that are not easily removed cannot be used, because they tend to build up on the hair, making it overconditioned and limp. The most widely used quaternary is stearalkonium chloride [122-19-0], which has been used at 3–5% concentration in cream rinses for many years. More recently, many useful quaternaries have become available; some are listed in Table 15. Table 15 also includes some quaternary polymers that are not only substantive to hair but also possess hair-fixative properties. Despite the commercial success of many conditioning quaternaries, efforts to synthesize better performing derivatives continue.

Table 15. Hair-Conditioning and Polymeric Fixative Compounds^a

Material	CAS Registry Number
<i>Hair conditioners</i>	
disoyadimonium chloride	[61788-92-9]
hydroxyethyl cetyldimonium chloride	[24625-03-4]
stearalkonium chloride	[122-19-0]
quaternium 22	[51812-80-7]
quaternium 79 hydrolyzed milk protein ^b	
<i>Hair conditioners with fixative properties</i>	
polyquaternium 4 ^c	
polyquaternium 6	[26062-79-3]
polyquaternium 7	[26590-05-6]
polyquaternium 10	[53568-66-4]
polyquaternium 11 ^d	
polyquaternium 22	[53694-17-0]
<i>Hair-fixative polymers</i>	
polyvinylpyrrolidinone (PVP)	[9003-39-8]
shellac	[9000-59-3]
vinyl acetate–crotonic acid–vinyl neodecanoate copolymer	[55353-21-4]

^a Ref. 26 includes a more comprehensive listing.

^b Material is the reaction product of a fatty acid amide of *N,N*-dimethylpropylenediamine and epichlorohydrin and hydrolyzed milk protein.

^c Dialkyldimethyl ammonium chloride–hydroxyethylcellulose copolymer.

^d Vinylpyrrolidinone–dimethylaminoethyl methacrylate copolymer, dimethyl sulfate reaction product.

A third type of hair-conditioning product relies on the use of proteins, amino acids (qv), botanicals, and amphoterics. Many ingredients have been identified as hair conditioners. Some of them are claimed to be substantive to the hair, whereas others are claimed to penetrate into the hair and repair previously incurred damage. Some of these hair-conditioning substances have been incorporated into newer delivery systems, such as mousses. These water-based products are dispensed as foams, which are rubbed into the hair and may then be rinsed off with water or allowed to remain on the hair for conditioning and styling benefits. All types of hydrolyzed proteins, such as keratin, soy, yeast, and wheat, chemically modified and free amino acids, such as cystine, aspartic acid, and lauroyl glutamate, and biological additives, such as casein, beer, eggs, nettle extract, and horse chestnut extract, have been formulated into products containing amphoteric and other more conventional cosmetic ingredients. Reference 26 includes an extensive list of chemicals used in hair conditioners and related products.

Hair Fixatives. These products are designed to assist in hair styling and in maintaining the style for a period of time. In contrast with hair dressings, hair fixatives do not leave an oily residue on the hair but tend to coat the hair with film-forming residues after drying. As in the case of hair dressings, style-holding qualities depend primarily on fiber–fiber adhesion and to a minor extent on fiber coating. The products may be conveniently divided into two groups: those that are applied to damp or wet hair, hair-setting products, and those that are applied to hair after styling, hair sprays. Styling requires that hair be formed into and retain the desired configuration. Curlers of various designs provide a wavy style, whereas hot combing results in essentially straight hair.

Wave-setting products can be applied to wet hair and should not interfere with or delay drying. Such products commonly aid in wet styling of hair. After drying, these products are claimed to help retain the style, regardless of frequent combing or exposure to high humidity. Wave sets can be formulated with water-soluble polymeric substances or with polymers that show solubility only in hydro-alcoholic media. Some of the preferred hair-fixative polymers (see Table 15) are combined with lubricants or emollients and other excipients. The viscosity of these products can vary from that of a water-thin fluid to a rather firm gel. The set-holding polymer constitutes about 1–3% of the product, and the viscosity-increasing substance is commonly a cross-linked polyacrylate, for example, carbomer [9007-16-3].

Hair sprays are applied from aerosol cans or pumps to dried and styled hair. Hair-spray products containing little or no water are preferred, because the presence of significant levels of water tends to soften a preexisting style. In the past, hair sprays were alcoholic solutions of polymers that were propelled with fluorinated or chlorinated highly volatile solvents. The alcohol concentration was kept as low as possible to reduce excessive wetting. As a result, the fixative resins had to exhibit good solubility in the propellant blends. Environmental regulations today preclude the use of these propellant solvents. Thus higher levels of alcohols are now used and the propellants of choice are low concentrations of hydrocarbons.

Hair Colorants. Hair colorants are commonly divided into temporary, semipermanent, and permanent types. Decolorizing (bleaching) represents a fourth type of hair coloring (81) (see BLEACHING AGENTS).

Hair bleaching removes the pigment melanin from the hair shaft by oxidative destruction. Alkaline hydrogen peroxide is the agent of choice. Because hydrogen peroxide is unstable at elevated pH, it is frequently supplied in pure form (6–10%) and is combined at the time of use with an ammonia or an amine-containing product to provide approximately 3–6% H₂O₂ at a pH of about 8.5 to 9.5. Thickening is required in order to retain the blended oxidizing mixture on the hair. Thickeners and conditioning agents, for example, fatty alcohols and protein derivatives, can be formulated into the alkalizing component or, occasionally, into the hydrogen peroxide. Surfactants are required to assure that every hair fiber is thoroughly wet by the blended mixture. Bleaching by hydrogen peroxide is enhanced by the presence of a peroxydisulfate, such as potassium persulfate [7727-21-1]. Bleaching damages the hair by converting some cystine to cysteic acid. In addition, the high pH induces swelling and cuticular damage. These adverse effects are counteracted by conditioning after treatment or by including some protectants in the hair-bleach product.

Temporary hair colorants are removed from the hair by a single shampoo. Temporary hair colorants usually employ certified dyes that have little affinity for hair (see Table 9). They are incorporated into aqueous solutions, shampoos, or hair-setting products.

Semipermanent hair colorants employ dyes that are absorbed directly by the hair. These dyes add color to the preexisting (natural) hair color and are useful primarily for blending in gray fibers. These dyes may fade significantly owing to exposure to sunlight and also are gradually removed by shampooing. Dyes selected for this purpose should not stain the scalp or skin during application. Typically, temporary hair colorings are distributed as pourable lotions. Formulations may include alkanolamides, polymeric substances, fatty alcohols, thickeners, and conditioners commonly employed in all hair cosmetics. The

chemical nature of the dyes is highly diverse and varies among manufacturers. Solvents, carriers, or complex solubilizers may be required when pigments are used. The CTFA lists temporary hair dyes with other substances as "Color Additives—Hair Colorants" (26). A few key chemical types are identified in Table 16.

Table 16. Semipermanent Hair Dyes^a

Name	CAS Registry Number	CI Number	Chemical type
Pigment Violet 19	[1047-16-1]	46500	quinacridone
Pigment Yellow 13	[5102-83-0]	21100	diazo
Basic Violet 3	[548-62-9]	42555	triaryl methane
Basic Red 76	[68391-30-0]	12245	monoazo
Direct Red 80	[2610-10-8]	35780	tetraazo
Disperse Blue 1	[2475-45-8]	64500	anthraquinone
HC Blue 2	[33229-34-4]		nitro- <i>p</i> -phenylenediamine
HC Yellow 4	[52551-67-4]		nitroaniline

^a Ref. 26 includes a comprehensive listing.

Development of the desired shade depends to a large degree on the tendency of the individual dyes in the mixture to adhere to the hair in proper proportions. Performance evaluation on many different types of hair, for example, natural, bleached, and permanently waved, is required.

Permanent hair colorants, frequently identified as oxidation dyes, show much greater resistance to fading and shampoo loss than do semipermanent hair colorants. As a rule, these dyes remain on the hair; it is common practice to dye only that portion of the hair shaft that has emerged from the scalp since the last application. In permanent dyeing, a lotion containing developers and couplers is blended with hydrogen peroxide and then applied to the hair. The objective is uniform penetration of the various components into the hair, oxidation of the developer to a reactive intermediate, and formation of a colored dye stuff with the coupler. The dyes are synthesized within the fiber and migrate outward only slowly because of their size. In addition, the reactions occur in alkaline media, and some of the peroxide bleaches the hair. Thus it is possible to generate colored hair lighter than the original shade. Dye formation is complex because shading is achieved by employing several developers and several couplers in the same dye bath. The process is illustrated by *p*-phenylenediamine, which is oxidized by the peroxide to a quinone diimine. This short-lived intermediate can react, for example, with resorcinol to yield a brownish indoaniline. Table 17 provides some insight into the many interactions that exist from just a few components. Further shading is possible by including semipermanent colorants (see Table 16), especially nitroaniline derivatives.

Table 17. Intermediates Used in Oxidation Hair Dyes^a

Material	CAS Registry Number	Molecular formula
<i>Developers</i>		
4-amino-3-nitrophenol	[119-34-6]	C ₇ H ₅ N ₂ O ₃
<i>p</i> -aminophenol	[123-30-8]	C ₇ H ₇ NO
2-methoxy- <i>p</i> -phenylenediamine sulfate	[42909-29-5]	C ₇ H ₁₀ N ₂ O ·xH ₂ SO ₄
<i>p</i> -methylaminophenol	[150-75-4]	C ₇ H ₉ NO
<i>p</i> -phenylenediamine	[106-50-3]	C ₇ H ₈ N ₂
<i>N</i> -phenyl- <i>p</i> -phenylenediamine	[101-54-2]	C ₁₂ H ₁₂ N ₂
phloroglucinol	[106-73-6]	C ₉ H ₆ O ₃
toluene-2,5-diamine	[95-70-5]	C ₇ H ₁₀ N ₂
<i>Couplers</i>		
<i>o</i> -aminophenol	[95-55-6]	C ₇ H ₇ NO
2,4-diaminophenoxyethanol HCl	[66422-95-5]	C ₈ H ₁₂ N ₂ O ₂ ·2HCl
2,6-diaminopyridine	[141-86-6]	C ₅ H ₇ N ₃
hydroquinone	[123-31-9]	C ₆ H ₄ O ₂
1,6-naphthalenediol	[83-56-7]	C ₁₀ H ₈ O ₂
<i>m</i> -phenylenediamine	[106-45-2]	C ₇ H ₈ N ₂
pyrocatechol	[120-80-9]	C ₆ H ₄ O ₂
pyrogallol	[87-66-1]	C ₆ H ₃ O ₃
resorcinol	[108-46-3]	C ₆ H ₄ O ₂

^a Ref. 26 includes a comprehensive listing.

In hair coloring a light ash blond shade may require as little as 0.5–1% of intermediates, whereas a true black may require up to about 5%. In principle, the formulator blends precursors that yield red, blue, and yellow dyes. The base in which the components are dissolved or suspended is similar to that used in simple bleaches and may include alkanolamides, various types of surfactants, thickening agents, and solvents. Removal of undesirable dyes is achieved by treating the discolored hair with a powerful reductant of the sulfite family.

Permanent coloration can also be achieved by exposing hair to certain metals: copper, silver, and especially lead salts. Preparations containing aqueous solutions of lead acetate may include a source of sulfur, usually thiosulfate, which may react with cystine in the hair to produce some cysteine or may react directly with the metal ion to form dark metallic sulfides. Preparations of this type, which darken hair gradually, are not universally considered safe.

Hair Waving and Straightening Products. The development of hair-waving and hair-straightening products requires a careful balance between product performance and hair damage. The hair-waving process essentially depends on converting some cystine cross-links in keratin to cysteine residues, which are reoxidized after the configuration of the hair has been changed. Sometimes hair straightening can also be achieved by a similar, relatively innocuous chemical change in the hair. As a rule, however, much more chemical destruction is required to achieve rapid and permanent straightening than to achieve permanent waving.

Permanent waving depends on the metathesis of a mercaptan and the cystine in hair while the hair is held in a curly pattern on a suitable device (cudrer). The most commonly used mercaptan is thioglycolic acid [68-11-1], although some other nonvolatile mercaptans can be employed. The active species is the mercaptide anion. Thus, adjustment to a pH between about 8.8 and 9.5 using amines or especially ammonia is required. A typical hair-waving product may consist of a 0.5–0.75 *N* solution of thioglycolic acid adjusted to a pH of about 9.1 with ammonia. The product generally includes a nonionic surfactant, to ensure thorough wetting of the wound hair tress, and a fragrance. Opaque lotion products can be created by adding the actives to mineral oil or other lipid-containing emulsions. The thioglycolate lotion is allowed to remain on the hair for about 10–30 min; then the hair is rinsed with water. Next,

an oxidizing solution consisting of a dilute (1.5–3°) acidified hydrogen peroxide solution or of a potassium or sodium bromate is applied to the hair. This so-called neutralizing solution oxidizes the cysteine residues to cystine (without bleaching) in a new configuration within about 5–10 min. The neutralizer may contain a variety of hair conditioners and is removed from the hair by thorough rinsing with water after unwinding.

Alternatively, the metathesis can be effected by sulfites or bisulfites that convert cystine into one cysteine residue and one thiosulfate (Bunte salt) residue. Hair waving based on sulfites is slower than that based on mercaptans and is more likely to cause changes in hair color.

Acidic waving systems based on mercaptans have recently achieved some popularity. The preferred mercaptan is glyceryl thioglycolate [30618-84-9], which is relatively odorless and provides sufficient active anionic mercaptide species at a neutral pH.

Hair straightening is more difficult than hair waving. Kinky hair has a tight crimp that cannot be straightened by winding over a rod or curler. Two processes for straightening exist. One, based on thioglycolates, effects the same chemical change as that occurring during permanent waving. The other, more aggressive, process is based on (1–8°) sodium hydroxide (or guanidine). The exact concentration depends on the temperature at which the process is carried out. In order to hold the hair straight, hair-straightening products are viscous. The hair is combed repeatedly during the process, which has the effect of reconfiguring the hair but can lead to serious hair damage from excessive pulling. The chemical reactions with sodium hydroxide involve formation of cysteine and dehydroalanine residues in the hair with some loss of sulfur. The cysteine and dehydroalanine can subsequently react to form the thioether, lanthionine [922-55-4], which helps repair the mechanical strength of the fiber to some extent. Similar chemical reactions occur when steam is allowed to interact with hair, such as during hot pressing, which was an earlier technique for straightening hair.

Conditioners, lipids, acid rinses, and related cosmetics have been developed to minimize hair damage from these rather destructive processes.

Hair Removers. Hair removers are designed to remove hair from the skin surface without cutting in order to avoid undesirable stubble. Cosmetic products have been developed for chemical destruction of hair, that is, depilation, and for facilitating mechanical hair removal, that is, epilation. Depilatories epitomize the chemical destruction of hair and allow hair removal by scraping with a blunt instrument or by rubbing with terry cloth. Chemical depilatories are based on 5–6° calcium thioglycolate in a cream base (to avoid runoff) at a pH of about 12. The pH is maintained with calcium or strontium hydroxide. Hair destruction is rapid, requiring not more than about 10 min. Treatment with a depilatory is followed by careful rinsing with water and various conditioning products intended to restore the skin's pH to normal. This type of treatment does not destroy the dermal papilla, and the hair grows back.

Epilation is required for permanent hair removal. The most effective epilation process is electrolysis or a similar procedure. Epilation can also be achieved by pulling the fibers out of the skin. For this purpose, wax mixtures (rosin and beeswax) are blended with lipids, for example, oleyl oleate, which melt at a suitable temperature (about 50–55°C). The mixture is applied to the site (a cloth tape may be melted into the mass) and after cooling is rapidly pulled off the skin. A similar process can be carried out with a tape impregnated with an aggressive adhesive.

Compositions have been published for cream rinses (82), hair conditioners, dressings, and mousses (83), hair-styling products (84), hair sprays (85), hair colorants (86), hair-waving products (87), hair-straightening products (88), and depilatories (89).

Decorative Cosmetics

Decorative cosmetics are products intended to enhance appearance by adding color or by hiding or deemphasizing physical defects. In Western cultures, most decorative cosmetics are for use on the face. Products in this category are various types of powders, facial makeups, and lip- and eye-coloring products. Regardless of the site of application or the type of product, all decorative cosmetics must meet certain critical performance criteria as outlined in Table 18. The dyes and pigments used in these products must be stable in the finished preparation and must not fade or discolor as a result of exposure to the variable environment of the skin.

Table 18. Performance Criteria for Decorative Cosmetics

Characteristic	Typical components	Comment
covering power	titanium dioxide zinc oxide magnesium oxide zirconium silicate	hides defects
slip	talc zinc or magnesium stearate starch emollient lipids	easy application smooth sensation
absorbency	chalk silica starch	absorbs skin secretions without color change
adherence	synthetic polymers emollient lipids volatile solvents polymeric substances gums	clings to skin limits ruboff dries to hard film
pigmentation	certified pigments noncertified pigments	provides color

Lip Makeups. Intensely pigmented coloring products have been used for many years to accentuate and modify the appearance of the lips. These products are marketed in soft stick forms (lipstick), as pastes (tinted lip gloss), and as hard sticks (lip liner). Lipsticks are manufactured via the molding process. The brilliant colors required for lipsticks are produced primarily by a limited number of available organic dyes and lakes. Formulation of acceptable shades is difficult and subject to the vagaries of fashion. For many years, lipsticks that caused a permanent stain on the lips were popular. Such staining is no longer desirable, and skin-staining dyes such as eosin (D&C Red No. 21) are rarely used. Traces of an inoffensive fragrance and of an antioxidant are commonly included in the lipid base (see Table 10). Sophisticated lipsticks may also contain moisturizers and uv light screens.

High gloss lipsticks use castor oil or 2-octyldodecanol. The more wear-resistant fat-based sticks are generally somewhat duller. There is a large number of possible ingredients, but the performance of most sticks is comparable. Feathering of the lipstick film, that is, creeping of color into crevices surrounding the lip tissue, is avoided by controlling the rheological properties of the applied stick mass.

Tinted and untinted soft lipstick masses are distributed in pans as lip glosses. These are applied with the fingers or with lipstick brushes. Hard, tinted, pencil-type sticks have been marketed as lip liners. Chap sticks are unpigmented lipsticks intended to alleviate scaling and to prevent cracked lips. Compositions of lipsticks, lip glosses, and chap sticks have been published (90).

Eye Makeup. Since antiquity, eye makeup preparations have been used to beautify the area surrounding the eye. Various forms of eye shadows color the eyelid; mascaras color and lengthen the eyelashes; eyeliners delineate the portion of the eyelid from which the eyelashes emerge. The appearance of eyebrows can be altered with various types of makeup, and, finally, false eyelashes and eyebrows have been marketed for years.

In the United States the use of coal-tar dyes in eye makeup is generally prohibited. The use of permanent and temporary hair colorants (Tables 16 and 17) and of organic dyes and their lakes is precluded. As a result, only insoluble inorganic pigments can be used (Table 9). The sensitivity of the eye

mandates that coarse or irritating particulate matter not be used. All eye preparations must be properly preserved and, preferably, self-sterilizing to prevent accidental introduction of pathogens into the eye.

Eye shadows may be molded or compressed. The technology used for molded sticks resembles that used in lipsticks except that the choice of pigments is restricted. Glossy lipids are generally avoided. If gloss is desired, it is achieved by the inclusion of a nacreous pigment. The basic technology for producing compressed eye-shadow sticks or powder compacts is that of other powder products.

Mascaras are available in three basic forms: cakes; creams; and mascaramatics, narrow containers provided with brushes. Cakes are commonly prepared by milling pigments into (sodium) soap chips and compressing this blend into metal or plastic pans. A small brush wet with water is used to transfer the mascara to the eyelashes. Cake mascaras can be modified with ingredients that improve adhesion of the mascara to the eyelashes, for example beeswax [8006-40-4] and dihydroabietyl alcohol [26216-77-31]; resist washing off or smearing by tearing, for example, aluminum mono-, di-, or tri-stearates; and lengthen the hairs (polymeric fibrous filaments).

Cream mascaras are pigmented, viscous, oil emulsions; soap emulsions are common. Viscosity is increased with glyceryl monostearate to help suspend the pigment. Cream mascaras are distributed in small jars or narrow-orifice tubes and are applied with brushes.

Mascaramatic mascaras have the largest share of the market. Emulsion mascaramatics are cream-type mascaras dispensed from containers that include a closure provided with a wand ending in a small brush. In solvent mascaramatics, mascara masses are pigment suspensions in thickened hydrocarbon solvents such as isoparaffins and petroleum distillates. The thickeners include waxes (microcrystalline [63231-60-7], carnauba [8015-86-9], or ouricury [68917-70-4], polymers (hydrogenated polyisobutene [61693-08-1]), and esters (propylene glycol distearate [6182-11-2] or trilaurin [538-24-9]).

Mascara pigmentation is usually black or brown-black. Mascaras during and after application are extremely close to the cornea, and any potential irritant must be rigidly excluded. The use of lash-elongating synthetics, such as rayon, nylon, and the like, has not resulted in significant problems.

The use of false eyelashes is rare. These are prepared from natural or synthetic fibers attached to a tinted lash strip, which can be glued onto the lid with the aid of an adhesive.

Eyeliners are available in two popular forms. One of these is a deeply pigmented emulsion that is applied with a fine brush. The emulsion must be viscous to avoid running and should dry to a waterproof film. The emulsion can be patterned after the emulsions used in mascaras. Glossy eyeliners require the use of nacreous pigments suspended in polymeric film formers, for example, acrylic acid copolymers.

The second type of eyeliner is a soft crayon, in pencil form, that delivers mass with minimal pressure. Products of this type may contain as much as 70% talc, 5% pigment, and 5% aluminum stearate. The lipid portion may include squalene [111-02-4] and alkanolamides.

Eyebrow pencils are used to outline the brow, especially after plucking of undesirable fibers. They are commonly prepared by stick molding. Eyebrow pencils are generally harder than eyeliner pencils. Like other pencils, the extruded mass is normally encased in wood. Some eyebrow pencils may be quite soft, approaching the texture of lipsticks.

Formulations of eye shadows (91), mascaras (92), and eyeliners (93) have been published.

BIBLIOGRAPHY

"Cosmetics" in *ECT* 1st ed., pp. 545–562 by F. E. Wall, Consulting Chemist; in *ECT* 2nd ed., Vol. 6, pp. 346–375 by H. Isacoff, International Flavors and Fragrances (U.S.); in *ECT* 3rd ed., Vol. 7, pp. 143–176, by H. Isacoff, International Flavors and Fragrances, Inc.

1. H. Truttwin, *Handbuch der Kosmetischen Chemie*, 2nd ed., J. A. Barth, Leipzig, Germany, 1924.
2. M. G. de Navarre, *The Chemistry and Manufacture of Cosmetics*, D. Van Nostrand Company, Inc., New York, 1941; F. Chilson, *Modern Cosmetics*, The Drug and Cosmetic Industry, New York, 1934 (formulation only).
3. *CTFA Buyer's Guide*, CTFA, Washington, D.C.
4. *Cosmetics & Toiletries, Who's Who*, Allured Publishing Co., Wheaton, Ill., Jan., 1993.
5. *Soap Cosmetics Chemical Specialties, Blue Book*, PTN Publishing Co., Melville, N.Y., Apr.
6. *CFR* 21, Part 211, U.S. Government Printing Office, Washington, D.C. 20402.
7. *CTFA International Cosmetic Ingredient Dictionary*, 4th ed., CTFA, Washington, D.C., 1991.
8. A. C. deGroot and I. R. White, *Contact Dermatitis* **25**, 273 (1991).
9. *CTFA International Resource Manual* and *CTFA International Color Handbook*, CTFA, Washington, D.C., 1992.
10. P. M. Silber, *Cosm. Toil.* **106**(IV), 56 (1991).
11. J. H. Draize and co-workers, *J. Pharmacol., Exp. Ther.* **82**, 377 (1944); *CFR* 16, 1500.42.
12. H. P. Frosch and A. M. Kligman, *Contact Dermatitis* **2**, 314 (1976).
13. H. P. Frosch, in H. P. Frosch, ed., *Principles of Cosmetics for Dermatologists*, Mosby, St. Louis, Mo., 1982.
14. B. Magnusson and A. M. Kligman, *J. Invest. Dermatol.* **52**, 268 (1969).
15. A. M. Kligman, *J. Invest. Dermatol.* **47**, 393 (1966).
16. P. D. Forbes and co-workers, *Food Cosm. Toxicol.* **15**, 55 (1977).
17. T. Maurer, *Contact and Photocontact Allergens*, Marcel Dekker, New York, 1983.
18. L. C. Harber and A. R. Shalita, in F. N. Marzulli and H. I. Maibach, eds., *Dermatotoxicology and Pharmacology*, Hemisphere, Washington, D.C., 1977.
19. K. H. Kaidbey and A. M. Kligman, *Contact Dermatitis* **6**, 161 (1980).
20. J. H. Draize and E. A. Kelley, *Proc. Sci. Sect. Toilet Goods Assoc.* **17**, 1 (1952); *CFR* 16, 1500.41.
21. A. M. Kligman and O. H. Mills, Jr., *Arch. Dermatol.* **106**, 843 (1972).
22. H. P. Frosch and A. M. Kligman, *J. Soc. Cosm. Chem.* **28**, 197 (1977).
23. K. Lammintausta and co-workers, *Dermatosen* **36**, 45 (1988).
24. M. Lanzet, *Cosm. Toil.* **101**(II), 63 (1986); *Cosm. Toil.* **105**(X), 26 (1990).
25. *CTFA Microbiology, Quality Assurance, and Occupational Safety Guidelines*, CTFA, Washington, D.C.; *Cosm. Toil.* **105**(III), 79 (1990).
26. *CTFA International Cosmetic Ingredient Handbook*, 2nd ed., CTFA, Washington, D.C., 1992.
27. D. S. Orth, *Cosm. Toil.* **106**(III), 45 (1991).
28. W. J. L. Smith, in N. Estrin, ed., *The Cosmetic Industry*, Marcel Dekker, New York, 1984, Chapt. 21.
29. S. Friberg, *J. Soc. Cosm. Chem.* **30**, 309 (1979).
30. P. Becher, *J. Disp. Sci. Technol.* **5**, 81 (1984).
31. Y. Saiato and co-workers, *JAOCs* **67**, 145–148 (1990).
32. B. Idson, in H. A. Lieberman and co-eds., *Pharmaceutical Dosage Forms, Disperse Systems*, Vol. I, Marcel Dekker, New York, 1988, Chapt. 6.
33. M. Rosoff, in Ref. 32, Chapt. 7.
34. K. Shinoda and H. Kunieda, in P. Becher, ed., *Encyclopedia of Emulsion Technology*, Vol. 1, M. Dekker, New York, 1983.
35. M. M. Rieger, in L. Lachman and co-eds., *The Theory and Practice of Industrial Pharmacy*, 3rd ed., Lea and Febiger, Philadelphia, 1986.
36. M. M. Rieger, *Cosm. Toil.* **106**(V), 59–69 (1991).
37. H. E. Junginger, *Pharm. Weekblad* **6**, 141 (1984).
38. G. M. Eccleston, *J. Soc. Cosm. Chem.* **41**, 1–22 (1990).
39. C. Prybiski and co-workers, *Cosm. Toil.* **106**(XI), 97 (1991).
40. A. Bevacqua and co-workers, *Cosm. Toil.* **106**(V), 53 (1991).
41. F. Comalles and co-workers, *Int. J. Cosm. Sci.* **11**, 5 (1989).
42. M. Riaz and co-workers, in H. H. Lieberman, ed., *Pharmaceutical Dosage Forms: Disperse Systems*, Vol. 2, Marcel Dekker, New York, 1989, Chapt. 16;

N. Wiener, *Antimicr. Agents Chemother.* **34** (10), 107 (1990).

43. J. N. Israelachvili, *Intermolecular and Surface Forces*, Academic Press, New York, 1985.

44. A. J. I. Ward and C. duReau, *Int. J. Pharm.* **74**, 137 (1991).

45. R. O. Potts and co-workers, *J. Invest. Dermatol.* **96**, 495, 580 (1991).

46. G. Imokawa and co-workers, *J. Invest. Dermatol.* **96**, 529 (1991).

47. G. V. Civile and C. A. Dus, *Cosm. Toil.* **106**(V), 83 (1991).

48. *Cosm. Toil.* **107**(III), 96, 97 (1992); **106**(XII), 71 (1991).

49. *Cosm. Toil.* **106**(XII), 101–104 (1991); **107**(VII), 101–108 (1992).

50. *Cosm. Toil.* **106**(XII), 105, 106 (1991).

51. *Cosm. Toil.* **106**(XII), 97–100 (1991); **107**(VII), 110–114 (1992).

52. *Fed. Reg.* **56**, 41008–41020 (Aug. 16, 1991).

53. *Cosm. Toil.* **107**(III), 96 (1992).

54. *Fed. Reg.* **43**(166), 38206–38267 (Aug. 25, 1978).

55. N. A. Shaath, *Cosm. Toil.* **101**(III), 55 (1986); **102**(III), 21 (1987); **102**(III), 69 (1987); idem, in N. J. Lowe and N. A. Shaath, eds., *Development, Evaluation and Regulatory Aspects*, Marcel Dekker, New York, 1990; *Fed. Reg.* **43**(166); 38206–38269 (Aug. 25, 1978).

56. *Cosm. Toil.* **105**(XII), 92, 122–130, 133–134 (1990); **106**(XII), 107, 108 (1991).

57. *Cosm. Toil.* **104**(VII), 83–84 (1989).

58. *Cosm. Toil.* **104**(VII), 85 (1989).

59. *Cosm. Toil.* **104**(VII), 86–94 (1989); **106**(XII), 109–110 (1991).

60. *Cosm. Toil.* **107**(V), 24 (1992); **106**(IX), 32 (1991).

61. *Cosm. Toil.* **106**(VI), 33, 34 (1991); **106**(XII), 107, 108 (1991).

62. *Fed. Reg.* **84**, 13440–13499 (Apr. 3, 1989).

63. *Fed. Reg.* **47** (162), 36492–36505 (1982).

64. *Cosm. Toil.* **102**(X), 54 (1987); **105**(IV), 76–80(1990); **106**(XII), 78, 110 (1991).

65. *Cosm. Toil.* **102**(X), 53, 119 (1987); **105**(IV), 75, 76 (1990); **106**(XII), 78 (1991).

66. **U.S. Pat. 4,954,282** (Sept. 4, 1990), K. J. Rys and co-workers (to Lever Brothers Co.).

67. *Fed. Reg.* **56**, 63554–63569 (Dec. 4, 1991).

68. *Cosm. Toil.* **102**(X), 112, 113, 116–117 (1987); **106**(VI), 34 (1991).

69. *Cosm. Toil.* **102**(X), 115 (1987); **106**(IV), 89–93 (1991).

70. *Cosm. Toil.* **106**(IV), 83–86 (1991); **107**(II), 29 (1992).

71. *Cosm. Toil.* **106**(IV), 91–92 (1991).

72. *Cosm. Toil.* **107**(VII), 116, (1992).

73. *Cosm. Toil.* **102**(X), 87 (1987); **107**(III), 93–95 (1992).

74. *Cosm. Toil.* **107**(III), 95 (1992).

75. *Cosm. Toil.* **105**(IV), 81 (1990).

76. *Cosm. Toil.* **105**(IV), 81–84 (1990).

77. *Cosm. Toil.* **105**(IV), 84–87 (1990); **106**(XII), 111 (1991).

78. L. J. Viola, in E. Sagarin and co-eds., *Cosmetics: Science and Technology*, 2nd ed., Vol. 2, John Wiley & Sons, Inc., New York, 1972, p. 543.

79. *Cosm. Toil.* **106**(VI), 36 (1991).

80. *Cosm. Toil.* **104**(VII), 97, 98 (1989).

81. J. F. Corbett, *Cosm. Toil.* **106**(VII), 53 (1991).

82. *Cosm. Toil.* **106**(IV), 94–98 (1991).

83. *Cosm. Toil.* **102**(X), 107–109, 110–111 (1987); **106**(VII), 72, 73, 75, 76, 83 (1991).

84. *Cosm. Toil.* **106**(VII), 78–80 (1991).

85. *Cosm. Toil.* **106**(VII), 29, 80, 82 (1991).

86. *Cosm. Toil.* **106**(VII), 71 (1991).

87. *Cosm. Toil.* **102**(X), 109–110 (1987); 106 (VII), 77 (1991).

88. *Cosm. Toil.* **106**(VII), 49, 77, (1991).

89. *Cosm. Toil.* **105**, 87 (1990).

90. *Cosm. Toil.* **104**(VII), 67–76 (1989).

91. *Cosm. Toil.* **104**(VII), 76–78 (1989).

92. *Cosm. Toil.* **104**(VII), 79–81 (1989); **106**(IX), 33 (1991).

93. *Cosm. Toil.* **104**(VII), 78–79 (1989); **106**(VI), 36 (1991).

General References

E. Jungermann, ed., *Cosmetic Science and Technology Series*, Marcel Dekker, New York, continuing series.

Kosmetikjahrbuch, H. Ziolkowsky KG, Augsburg, Germany.

F. N. Marzulli and H. I. Maibach, eds., *Dermatotoxicology*, 4th ed., Hemisphere Publishing Co., Washington, D.C., 1991.

A. Nowak, *Cosmetic Preparations*, Micelle Press, London, 1991.

M. Rieger, *Surfactants in Cosmetics*, Marcel Dekker, New York, 1985.

K. Schrader, *Grundlagen und Rezepturen der Kosmetika*, H thtig Verlag, Heidelberg, Germany, 1989.

W. C. Waggoner, ed., *Chemical Safety and Efficacy Testing of Cosmetics*, Marcel Dekker, New York, 1990.

Technical Journals

 rztliche Kosmetologie, G. Braun, Karlsruhe, Germany.

International Journal Cosmetic Science, Blackwell Scientific Publications, Oxford, UK.

Japanese Cosmetic Science Society Journal, Tokyo, Japan.

Journal of the Society of Cosmetic Chemists, New York.

Trade Journals

Cosmetics & Toiletries, Allured Publishing Co., Wheaton, Ill.

Drug & Cosmetics Industry, Edgell Communications, Inc., Cleveland, Ohio.

Fett-Wissenschaft Technologie, Konradin-Industrieverlag, Leinfelden-Echterdingen, Germany.

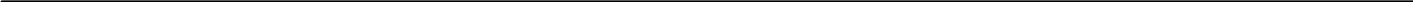
HLAPPI, Rodman Publishing Corp., Ramsey, N.Y.

Manufacturing Chemist, Morgan-Grampian, London.

Parf merie und Kosmetik, A. H thtig, Heidelberg, Germany.

Parfums, Cosmetiques, Aromes, Paris.
Seifen, Oele, Fette, Wachse, A. Ziolkowsky KG, Augsburg, Germany.
Soap Cosmetics Chemical Specialties, PTN Publishing Co., Melville, N.Y.
Soap, Perfumery and Cosmetics, United Trade Press, Ltd., London.

Martin M. Rieger
M & A Rieger, Associates



COTTON

The use of cotton predates recorded history. Although the actual origin of cotton is still unknown, there is evidence that it existed in Egypt as early as 12,000 BC. Archaeological findings indicate its use in cloth in 3,000 BC, and there are records of its cultivation in India as far back as 700 BC. In the fifth century BC, Herodotus wrote of "trees" growing wild in India bearing wool of a softness and beauty equivalent to that of the sheep; clothes made from this tree wool were described as garments of extraordinary perfection.

It is said that Alexander the Great introduced Indian cotton into Egypt in the fourth century BC, and from there it spread to Greece, Italy, and Spain. During the year AD 700, China began growing cotton as a decorative plant, and AD 798 saw its introduction into Japan. Early explorers in Peru found cotton cloth on exhumed mummies that dated to 200 BC. Cotton was found in North America by Columbus in 1492. About 300 years later, the first cotton mill was built in Beverly, Massachusetts, and in 1794 Eli Whitney was granted a patent for the invention of the cotton gin.

Cotton is the most important vegetable fiber used in spinning (see FIBERS, VEGETABLE). Its origin, breeding, morphology, and chemistry have been described in many publications (1–4). It is a member of the Malvaceae, or mallow family, is a plant of the genus *Gossypium*, and is widely grown in warm climates the world over.

The average cotton plant is a herbaceous shrub with a normal height of 1.2 to 1.8 m, although some tree varieties reach a maximum height of 4.5 to 6.0 m. The most important species included in the genus *Gossypium* are *hirsutum*, *barbadense*, *aboreum*, and *herbaceum*.

G. hirsutum originated in Central America and was part of the Mayan culture of Mexico. This species comprises all of the many varieties of American Upland cotton of commerce. *G. hirsutum*, a shrubby plant that reaches a maximum height of 1.8 m, is probably the origin of the green-seeded (ie, bearing green fuzz fibers) cotton of former times. *G. hirsutum* is grown extensively in the southern United States.

G. barbadense, originally from the Incan civilization, grows from black seeds and reaches a height of 1.8 to 4.5 m. It is the species with the longest staple and includes Sea Island, Egyptian Giza strains, American–Egyptian, and Tanguis cotton. Sea Island is the longest and silkiest of the commercial cottons.

G. arboreum, the tree wool of India, grows as tall as 4.5 to 6.0 m and includes both Indian and Asiatic varieties. Its seeds are covered with greenish gray fuzz fibers below the white lint fibers.

G. herbaceum, the original cotton of India, averages 1.2 to 1.8 m in height. The fiber is grayish white and grows from a seed encased in gray fuzz fibers. This species includes the short staple cottons of Asia and some of China as well as most native Indian varieties. It is also grown commercially in Iran, Iraq, Turkey, and the former USSR.

The most favorable growing conditions for cotton include a warm climate (21 to 30°C, mean temperature) for at least 150 days. Fairly moist and loamy soil produces the highest yields. Under normal climatic conditions, cotton seeds germinate in 7 to 10 days. Flower buds (known as squares) appear in 35 to 45 days followed by open flowers 21 to 25 days later. One day after the flower opens the cotton boll begins to grow rapidly, if the flower has been fertilized. The mature boll opens 45 to 90 days after flowering, depending on variety and environmental conditions. Within the boll are three to five divisions called locks, each of which normally has seven to nine seeds that are covered with both lint and fuzz fibers (Fig. 1). The fuzz fibers form a short, shrubby undergrowth beneath the lint hairs on the seed. At least 10,000 fibers are attached to each seed and there are close to 500,000 fibers in each boll. The usual range of planting time in the United States extends from the beginning of March to the end of May; harvest time is in the late summer or early fall.

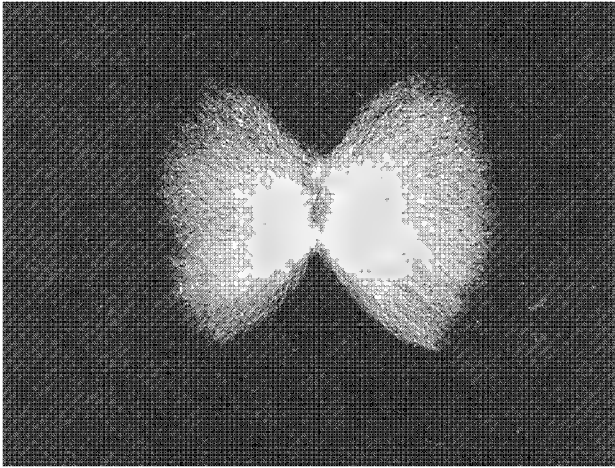


Fig. 1. Cotton butterfly with lint and fuzz fibers.

The cotton fiber is a single cell that originates in the epidermis of the seed coat at about the time the flower opens; it first emerges on the broad, or chalazal, end of the seed and progresses by degrees to the sharp, or micropylar, end. As the boll matures, the fiber grows until it attains its maximum length, which averages about 2500 times its width (Fig. 2). During the first three weeks, the cell is composed of only a thin wall that is covered with a waxy, pectinaceous material, which encloses the protoplasm or plant juices. In about 17 to 25 days after the flower opens, when the boll is half mature, each fiber virtually attains its full length. Then layers of cellulose (qv) are deposited on the inside of the thin casing, or primary wall. The pattern of deposition is such that one layer of cellulose is formed each day in a centripetal manner until the mature fiber has developed a thick secondary wall of cellulose from the primary wall to the lumen, or central canal. The fiber now consists of three main parts: primary wall, secondary wall, and lumen. At the end of the growing period when the boll bursts open, the fibers dry out and collapse, forming shriveled, twisted, flattened tubes.



Fig. 2. Single cotton fibers, showing ratio of length to width.

The seed hairs of cultivated cottons are divided into two groups (fuzz and lint) that differ in length, width, pigmentation, and strength of adherence to the seed. The growth of fuzz fibers is much the same as that of lint, but they are usually about 0.33 cm long compared with the 2.5 cm average length of lint fibers and are twice as thick, or about 32 μm (Fig. 3). Their color ranges from greenish brown to gray. After lint fibers have been ginned off the seed, the fuzz fibers (which are also called linters) remain. Removal of linters requires a machine similar to the regular cotton gin. Linters are an important source of cellulose for chemical purposes and are used in upholstery and batting.



Fig. 3. Longitudinal view of fuzz (a) and lint (b) fibers.

Cultivation and Production

At present, the chief cotton-growing countries of the world are the People's Republic of China (23° o), the United States (17° o), the former USSR (15° o), India (11° o), Pakistan (8° o), Brazil (4° o), Turkey (3° o), and Egypt (2° o) (5). In the Western Hemisphere, cotton is cultivated between about 37° N and 32° S latitude; in the Eastern Hemisphere, the limits extend to 47° N and 30° S.

In the United States management and harvesting practices have increased the yield so only about half as much land is needed today to produce the

same amount of cotton as in 1930. Cultivation of cotton differs markedly from one country to another, depending on methods of mechanization. Approximately 70% of the U.S. cotton is rain grown, but western states (California and Arizona) only grow irrigated cotton. The use of supplemental irrigation is increasing in Texas and the Mid-South states.

Through cotton-breeding research, the United States has developed varieties that today remain the leading fine-quality cottons. Two such varieties, Deltapine 50 and Paymaster HS26, were planted on 30% of the U.S. cotton acreage (6).

Fertilizers. Cotton is normally grown under intense production systems; many fields are planted in cotton year after year. On average in 1990 U.S. cotton was fertilized with 31 kg of nitrogen, 10 kg of P₂O₅, and 6.8 kg of K₂O (7). The type and concentration of fertilizer required for high yield depends on many factors such as soil type, previous fertilization rate, cropping system, and irrigation. Therefore, an efficient fertilization program must be based on results of soil and tissue tests and the yield desired from the crop.

Throughout the cotton-growing regions of the United States, the method of applying fertilizer must be tailored to the crop needs and the characteristics of the cropping system. In some production systems, fertilizer is applied during the seedbed preparation, whereas in other systems it may be applied at planting or after emergence. Combinations of preplant and postemergence applications are common, especially for nitrogen. Foliar application is relatively new. Dilute nitrogen and phosphorus solutions are sprayed on the foliage of the plant at various times during the season. This is an attempt to match fertilizer application to the weekly needs of the plant more closely (see FERTILIZER).

Pests and Insecticides. The most destructive pests of the cotton plant are the boll weevil and the bollworm–budworm complex. They are serious threats to the cotton industry in countries around the world. The boll weevil migrated from Mexico around 1892 and spread over the entire cotton belt within 30 years. The domestic cotton crop lost to the weevil is worth \$200 million a year. In addition, about \$75 million a year is spent for pesticides to control this destructive pest (8). Unfortunately, some insecticides used to control the weevil kill many beneficial insects. Among the undesired casualties are insects that help to control the bollworm and the tobacco budworm, pests that cause another \$200 million loss in cotton.

In 1916, calcium arsenate [7778-44-1] dusted by airplane was used to control the boll weevil; however, throughout many developments in effective insecticides, such as organophosphates, the boll weevils became resistant to poisons that were formerly effective (see INSECT CONTROL TECHNOLOGY).

In the mid-1970s, a test program to eradicate the boll weevil, carried out on an 8100-ha (81-km²) region of Mississippi, Alabama, and Louisiana, to plans for beltwide eradication of this devastating pest (8). The pest could be eliminated from the United States by the year 2015.

Another destructive insect is the pink bollworm, which overwinters as diapausing (hibernating) larvae in the soil. After feeding on the late-blooming bolls, the larvae drop to the ground and hibernate for the winter, emerging as adults in the spring to lay eggs on the early cotton blooms. The eggs hatch and the new larvae bore into the fresh cotton bolls, go through molting stages, bore their way out, and drop to the ground. Throughout the growing season, the cycle repeats itself, rendering useless vast numbers of cotton plants in a single field.

A technique to combat this pest has been developed by agricultural research scientists in cooperation with the Arizona Agricultural Experiment Station in Phoenix. The number of overwintering pink bollworms is limited by reduction of their food supply late in the year. Chemical growth regulators are used to prevent late boll formation; this has no effect on the existing bolls and causes no loss in harvesting, as most of the late bolls fail to mature anyway. Along with the chemical treatments, other eradication practices include early stalk shredding, early and deep tillage, and winter irrigation that drowns diapausing larvae (9). Other insects injurious to the cotton plant include aphids, leafhoppers, lygus bugs, mites, whiteflies, fleahoppers, thrips, cutworms, and leaf miners (10).

Harvesting. Except for the cotton gin, the introduction of the mechanical harvester has probably had a greater effect on cotton production than any other single event. Commercial mechanical harvesters were introduced into the United States after World War II. By 1955, about 23% of the cotton was mechanically harvested. That value had increased to 85% by 1965. In the early 1990s more than 99% of the U.S. cotton crop was mechanically harvested, although cotton was still hand harvested in some other countries.

When the cotton boll reaches full maturity, it begins to lose moisture and opens. As the boll opens, the drying fiber fluffs or expands outward. After the seed cotton (fuzz and lint) has dropped to a moisture content of about 12% it is ready for harvest. If the cotton is to be mechanically harvested, it is usually treated with a harvest-aid chemical classified as either a defoliant or desiccant. A defoliant induces abscission (shedding) of foliage. The removal of leaves helps to minimize the trash harvested with the mechanical harvester and promotes faster drying of early morning dew on the lint. Defoliants should not be applied until about 60% of the bolls are open and harvest should be delayed for 7 to 14 days after application. Desiccants (qv) are chemicals that induce rapid loss of water from the plant tissue and subsequent death of the tissue. The dead foliage remains attached to the plant. Harvest can begin in 3 to 5 days after application.

Once the plant is ready, the cotton is mechanically harvested with either a spindle picker or cotton stripper. The spindle picker selectively harvests seed cotton from open bolls. The unopened bolls are left on the plant and can be picked at a later date. The spindle picker uses a rotating tapered barbed spindle to remove the cotton from the bur (seed case). The seed cotton is wrapped around the spindle, pulled from the bur, removed from the spindle with a rubber doffer, and then transferred to a basket.

Two types of cotton strippers are currently in use in the United States. The finger-type stripper uses multiple fingers made from metal angles with the vee turned up and operating at a 15°–20° approach angle with the ground. The roll-type stripper uses two 7-inch-dia (1708-cm-dia) stripper rolls angled 30° with the ground and rotating in opposite directions. Each roll consists of three brushes and three paddles mounted in alternating sequence. Both types of strippers are efficient and can harvest up to 99% of the cotton from the plant. They are nonselective and remove not only the seed cotton but also the cracked and unopened bolls, the burs, and other foreign matter. The extra foreign matter requires additional cleaning at the gin.

Because the gin capacity is usually not sufficient to keep up with the harvesters, the harvested cotton is often stored in a compacted module and ginned at a later date. The type of storage or seed cotton processing may place additional constraints on the harvest process. If the seed cotton is to be placed in module storage, the cotton should not be harvested until the moisture content is 12% or less and the harvested seed cotton should be free of green plant material such as leaves and grass.

Ginning. Gin equipment is designed to remove foreign matter, moisture, and cottonseed from raw seed cotton. Typical types of gin equipment are cylinder cleaners, stick machines, and lint cleaners for cleaning; shelf-type driers for removing moisture; and gin stands for separating the fiber from the cottonseed. The gin stand (Fig. 4) is actually the only item of equipment required to gin cotton, the other equipment is for trash removal and drying. About 636 kg of seed cotton is required to produce a bale (~217 kg) of lint cotton from spindle harvested cotton. The remainder of the material consists of about 363 kg seed and 55 kg trash and moisture. Typical gins contain three or four individual gin stands, each rated at 4 to 6 bales per hour. However, some gins contain as many as eight gin stands that produce 12 to 15 bales per hour. The greatest number (30,948) of gins existed in the United States in 1902. The majority were on plantations, and they processed 10.6 million bales (2.3 × 10⁹ kg) of cotton (11). Since then the number of gins has declined, and the average number of bales processed per gin has increased. In 1990, a total of 1,536 active gins handled a crop of 15,064,000 bales (~3.3 × 10⁹ kg) for an average of 9,807 bales (2.1 × 10⁶ kg) per gin plant (12). The number of bales produced in the United States varies substantially from year to year, which places a severe financial burden on the ginning industry.

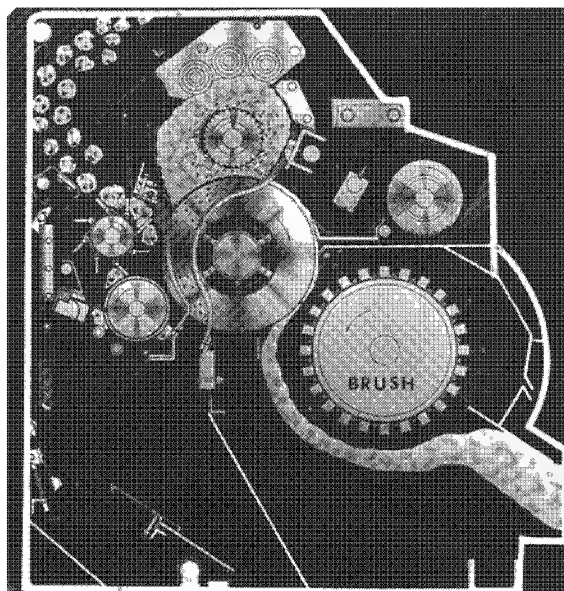


Fig. 4. A modern gin stand that separates fiber from cottonseed.

Mechanical harvesting systems were made possible by invention of saw-type lint cleaning systems in the early 1950s. Lint cleaners enabled gins to remove from the cotton the additional trash that resulted from mechanical harvesting. The mechanical systems reduced the harvesting period from 4 to 5 months to about 6 weeks of intensive operation. Severe congestion problems at the gin were eased by the advent of storage of seed cotton in 8- to 15-bale, freestanding modules. Modules avoided the massive need for wheeled trailers during the compressed harvest season. Storage of seed cotton in modules increased rapidly from the 1970s on and were used for more than one-half the crop in 1990. In 1957, the average U.S. cotton gin produced 20 bales (4340 kg) per hour; the present design capacity exceeds 30 bales per hour, and the average ginning capacity is about 15 bales per hour. Some gins process in excess of 100 bales per hour (13).

Most of the U.S. gins are now operated as cooperatives or as corporations serving many cotton producers. Automatic devices do the work faster, more efficiently, and more economically than hand labor. High volume bulk seed cotton handling systems and hydraulic suction systems to remove cotton from modules, high volume trailers to get cotton into the gin, larger trailers and modules, increased processing rates for gin equipment, automatic controls, automated bale packaging and handling devices, and improved management have all increased efficiency.

Greige Processing. The cotton from the gin is received by the textile mill in the form of highly compressed bales, weighing approximately 227 kg. Practically all the seed and a large portion of the other impurities are removed at the gin. However, the lint cotton still contains various forms of trash, including stem, leaf, and seed coat fragments that must be removed in the manufacturing process. Before it can be spun into yarn, the cotton must be opened, blended, cleaned, parallelized, and reduced to the proper size.

The first step in mill processing is opening and blending. Cotton properties vary considerably from bale to bale, depending on variety, growing conditions, and other production variables. For example, cottons grown under irrigation may have different properties and contain different levels of trash than those produced under natural rainfall conditions. Therefore, to ensure consistency in processing efficiency and product quality, it is important that many bales be blended together to produce a homogeneous mix. Most mills have their own formula for mixing cottons grown in various regions of the United States, depending on the quality and type of product produced. Sophisticated blending equipment in modern mills can mix more than 100 bales of cotton at a time.

After blending, the cotton is fed through a series of opening and cleaning machines containing various types of revolving beaters and sawtooth cylinders that reduce the cotton into smaller masses of less compacted tufts. Most of the dirt and heavier trash is removed through screens or grids as the cotton is tumbled, beaten, shaken, or otherwise manipulated.

The next step in the cleaning and opening process is carding. In this process the cotton is passed between two fine wire brushlike covered surfaces set in close proximity to each other and moving in opposite directions or in the same direction at different speeds. This results in a combing action and separates the fibers into a fine web. Because the cotton is in this opened condition, most of the remaining finer trash is removed. The fine web of fibers delivered from the card is condensed into a ropelike strand called card sliver and coiled into large cans.

Card sliver consists of a randomly oriented mass of fibers held together by the natural cohesiveness of the cotton. The fibers must now be parallelized longitudinally. In a process called drawing, several strands of card sliver (usually eight) are combined to produce a single sliver of improved uniformity and fiber orientation. The drawing frame contains four or five sets of drafting rolls rotating at progressively higher speeds, that attenuate or draft the material down to approximately the original size of the card sliver. The cotton sliver is generally processed through two or three drawings to obtain maximum uniformity and parallelization of individual fibers.

Sometimes cotton is combed to produce finer, higher quality yarns and fabrics. Combing mechanically removes many of the shorter cotton fibers from the stock. This is accomplished by firmly holding a portion of the parallelized fibers between nipper blades as a cylinder containing fine pins combs through the protruding fringe. The shorter fibers are removed by the pins and discharged separately by a brush and air suction. As much as 10 to 15% of the cotton is removed as short fiber in the combing process and finds use in the production of lower grades of fabric. Only about 12% of the cotton processed in the United States in 1990 was combed. Yarns not spun from combed cotton are referred to as carded yarns. Because many of the short fibers have been removed, combed yarns are stronger and more uniform than carded yarns. However, the combing process is expensive and adds considerably to yarn costs.

The next process, roving, is an intermediate step in the preparation of the cotton exclusively for ring spinning. On the roving frame, the sliver is attenuated several times by a series of drafting rollers and a small amount of twist is added to hold the smaller mass of stock together. The product, also called roving, is then wound on a special bobbin to accommodate the creel on a ring-spinning frame.

The final process in the yarn manufacturing operation is spinning. Ring spinning is the mainstay of the textile industry and accounts for over 50% of all cotton yarn produced in the United States. In ring spinning, the roving is first attenuated to the desired size through a series of drafting rollers. The strand of drafted fibers passes through a metal guide, or traveler, which revolves rapidly around a circular track, or ring, which in turn surrounds a rotating spindle and bobbin. Sufficient twist to obtain the required tensile strength is inserted by the rotation of the spindle and bobbin at speeds of up to approximately 20,000 rpm. Yarn is wound on the bobbin by an up and down traversing of the ring. Average production rates for ring-spun yarns range from 20 to 30 yards (18 to 27 m) per minute.

The newer spinning methods produce yarn directly from drawing sliver. Rotor, or open-end spinning is a relatively new method of yarn formation that can produce coarser yarns at three to five times the rate of ring spinning. Sliver is fed to a pinned or sawtooth-covered opening roller that rotates at a relatively high speed and individualizes the fibers. The opened fibers are then drawn via suction through a conical shaped duct where they are parallelized and deposited on the inside of a rapidly revolving rotor, from which they are twisted into yarn. Twist is inserted by rotation of the rotor, and the yarn is removed through a tube and wound on a package. Rotor-spun yarns are more uniform but weaker than ring-spun yarns.

Some other new methods of producing spun cotton yarns at exceptionally high production rates have been developed. Air-jet spinning is one of the newest yarn formation techniques and can spin yarns at speeds of up to 200 yards (183 m) per minute. A conventional roller drafting system is used to reduce drawing sliver to the proper size. The drafted ribbon of fibers is then opened, twisted, and entangled by jets of compressed air as it passes through nozzle assemblies. Air-jet spun yarns tend to be weaker and harsher than those produced by ring or rotor spinning. Another new method of yarn production in limited use is friction spinning. In this process the yarn is formed by frictional contact of the fibers with a pair of rotating perforated drums. The rotation of the drums causes the fibers to be rolled into a thread, which is then drawn off axially from the drums as a finished yarn. Production rates for this equipment can exceed 250 yards (229 m) per minute.

Most woven and knitted cotton fabrics are produced from single yarns as supplied by the spinning process. However, for the manufacture of industrial fabrics such as canvas, it is necessary to combine, or ply twist, several strands of single yarns together to obtain increased strength and resilience. Sewing thread and cordage are also produced from multiple plies of single yarns twisted together.

Physical Properties

Fiber length is universally accepted as the most important fiber property, because it greatly affects processing efficiency and yarn quality. The recognized reference machine method for fiber length information is the Suter–Webb Comb Sorter (14). Fibers are sorted and separated by a series of combs into length increments of 1.6 mm. Each group is then weighed to determine the weight–length distribution parameters, which include the mean length of the longest half (upper-half mean) of the fibers by weight, the mean length, the percentage of the fibers shorter than 12.7 mm and the coefficient of length variation. Variations of length are unique to specific varieties of cotton and range from less than 2.5 cm for short-staple Upland varieties to 2.6 to 2.8 cm for medium-staple Uplands, to >2.85 cm for long-staple varieties (Pima, Egyptian, and Sea Island) (15).

Next to length, fiber strength is the most important physical property that relates to fiber and yarn quality. The recognized reference method for fiber strength is based on measurements made on bundles of parallel fibers (16). One suitable instrument, the Pressley tester, consists of a set of jaws and an inclined lever system in which an ever-increasing load is applied to the specimen until the bundle breaks. The position of the load when the bundle breaks is read from the scale and used, together with the bundle weight, to calculate bundle strength. An alternate approach to bundle measurements is the Stelometer tester that uses a somewhat different loading concept but still requires clamped and weighed bundles very much like the Pressley. In this case, the clamp jaws are separated by 3.2 mm (gauge length), a convention that has proven the method to be highly related to yarn strength and processing parameters. Fiber strength is expressed as breaking stress or force to break per linear density of the bundle. These units are newtons (or gram force (gf)) per linear density (tex), where 1.0 tex = 1 g /1000 m. Variations in fiber strength are also unique to specific varieties of cotton and range from 0.176–0.216 N /tex (18 to 22 gf /tex) for short-staple Upland varieties to 0.235–0.275 N /tex (24 to 28 gf /tex) for some medium-staple Uplands to 0.314–0.373 N /tex (32 to 38 gf /tex) for long-staple varieties (Pima, Egyptian, and Sea Island) (17).

Another important characteristic property of cotton fibers is their fineness, or linear density, or weight per unit length. The normal units for fineness are millitex. Fineness is directly related to the amount of cellulose in the fiber, which is a function of the fiber wall area, excluding the hollow center (lumen), and the fiber length. Developments in computerized microscopic image analysis (18) allow for rapid and accurate measurements of fiber wall area and perimeter. The term *fiber maturity* relates to the degree of development or thickening of the fiber wall relative to its perimeter. Acceptable range of maturity for mill usage is from 75 to 80%. The most common measure of fiber fineness is the Micronaire reading, an airflow measurement performed on a 3.25-g test specimen, which is compressed to a specific volume in a porous chamber. Air is forced through the specimen and the resistance to the airflow is proportional to the linear density. The Micronaire reading is affected by a combination of both fiber fineness and maturity to the extent that for the same genetic variety with a constant perimeter, the Micronaire will directly correspond to maturity. Depending on acceptable maturity, a good range of Micronaire is between about 3.5 and 4.8 (19).

In addition to fiber length, strength, and fineness, two other properties that have significant bearing on fiber and yarn properties are color and trash measurements, which have been accomplished by instrumentation such as the Colorimeter and the Shirley Non-Lint Analyzer.

High Volume Instrument Systems. Instruments to measure fiber properties have been used for a number of years, but high costs and the length of time required for the tests have limited their use. However, in the mid-1960s, a cooperative effort between the USDA and instrument manufacturers began that was aimed at developing instruments for high volume use such as would be necessary for cotton classification. This led to the development of high volume instrument (HVI) systems. Modern HVI systems make use of the latest advances in electronic instrumentation and space-age technology to rapidly and inexpensively measure the more important fiber properties, including length, length uniformity, strength, fineness, color, and trash.

At present there are two HVI systems on the market. The first HVI system was developed by Motion Control Inc. (MCI) (Dallas, Texas) now owned by Peyer AG (Switzerland). The MCI system incorporates a distinctive clamping device controlled by air that grabs a tuft of fibers from the bulk sample held by the operator. The grab sample is held firmly by a clamp and is manually placed on an automatic comb and brusher, which straightens and aligns the beard of fibers before it is presented to the instrument for length measurement by an airflow method. The second type of HVI system was developed by Special Instruments Laboratory (Spinlab) (Knoxville, Tennessee) now owned by Zellweger Uster Corp. (Switzerland). Fiber beard samples for the Spinlab HVI are obtained with the cylindrically shaped Fibrosampler drum, which allows collection of fibers protruding through perforations in the drum into a special comb attached to the outside of the drum. The Spinlab instrument uses a light beam projected through the fiber beard to assess the amount of fiber in the cross section of the beard. Both instruments measure the force to break the respective fiber beards. The linear density of the broken fibers is estimated by airflow (MCI) and light attenuation (Spinlab) (20).

Currently, HVI systems are providing reliable information on six characteristics of quality from a cotton sample in approximately 30 s that are highly related to the spinning quality and market value of the cotton. As of the 1991 crop year, cotton must be tested by HVI to be eligible for price supports in the United States. Information on every bale of cotton should greatly improve the marketing of cotton and encourage the production of cotton with fiber properties desired by users.

Morphology

The cotton fiber is a single biological cell, 15 to 24 μm in width and 12 to 60 mm long; it has a central canal, or lumen, down its length except at the tip (21). It is tapered for a short length at the tip, and along its entire length the dried fiber is twisted frequently; the direction of twist reverses occasionally (22). These twists are referred to as convolutions, and it is believed that they are important in spinning because they contribute to the natural interlocking of fibers in a yarn.

Cotton is essentially 95% cellulose [9004-34-6] (Table 1). The noncellulosic materials, consisting mostly of waxes, pectinaceous substances, and nitrogenous matter, are located to a large extent in the primary wall, with small amounts in the lumen (23).

Table 1. Composition of Typical Cotton Fibers

Constituent	Composition, percent of dry weight	
	Typical	Range
cellulose	94.0	88.0–96.0
protein (% N × 6.25) ³	1.3	1.1–1.9
pectic substances	1.2	0.7–1.2
ash	1.2	0.7–1.6
wax	0.6	0.4–1.0

total sugars	0.3
pigment	trace
others	1.4

^a Standard method of estimating percent protein from nitrogen content (° N).

Of the noncellulosic substances in cotton, protein normally occurs in the largest amounts. This protein is apparently the protoplasmic residue left behind on the gradual drying up of the living cell. It occurs almost entirely in the lumen. On the assumption that the cell content is approximately the same for each fiber, it has been suggested that the nitrogen content might be used as an indirect measure of wall thickening. However, no practical test of sufficient accuracy has been developed.

Most of the pectin in the cotton fiber is in the primary wall. Removal of the pectic substances is accomplished by scouring, which does not change the properties of the cotton greatly.

The wax of most cottons is a soft, low melting complex mixture; there are some differences among varieties. The noncommercial, strongly pigmented green lint cotton contains about 17° wax of high melting point. Because the wax becomes established in fibers, largely if not wholly, during the first phase of development, the wax content expressed as a percentage of the whole fiber mass decreases as the fiber maturity or degree of wall thickening increases. Removal of cotton wax by chemical treatment increases the friction between fibers and between fiber and metal.

Although the mature cotton fiber is considered as having a primary wall, a secondary wall, and a lumen, further examination of its structure shows a cuticle and a winding layer (Fig. 5). The cellulose of the primary wall exists as a woven network of microfibrils in and on which are deposited noncellulosic materials that form the cuticle. Just beneath the primary wall is the winding layer, which is also the first layer of the secondary wall. The winding layer appears to be made up of a single layer of fibrillar bundles composed of microfibrils and oriented at an angle to the fiber axis. The main body of the fiber consists of cellulose fibrils packed tightly in a solid cylinder, which, under certain conditions of chemical swelling, can be induced to separate into more or less concentric layers. These layers seem to have a finer and more regular structure than does the winding layer, because the 20 to 50 secondary wall layers have cellulose microfibrils compactly parallelized along the axis of the fiber (see CELLULOSE).

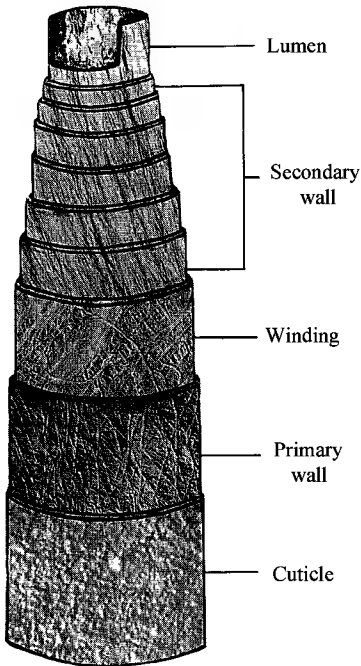


Fig. 5. Schematic diagram of cotton fiber.

The gross morphology of cotton, which refers to the relatively large structural elements above, is visible in the electron microscope. The microfibrillate structure includes pores, channels, and cavities that play an important role in the chemical modification of cotton. The fibrils follow a spiral pattern and at times reverse; it is believed that regions of low strength along the length of the fiber occur close to the reversal zones. Microfibrils of the secondary wall are 10 to 40 nm wide, and these in turn are composed of elementary fibrils (crystallites) 3 to 6 nm wide.

Chemical modification of the cotton fiber must be achieved within the physical framework of this rather complicated architecture. Uniformity of reaction and distribution of reaction products are inevitably influenced by rates of diffusion, swelling and shrinking of the whole fiber, and by distension or contraction of the fiber's individual structural elements during finishing processes.

Structure and Reactivity

Chemical Structure. The raw cotton fiber is predominantly cellulose, and its structures and chemical properties are essentially those of the cellulose polymer (see CELLULOSE). It is a linear condensation polymer of D-anhydroglucose units linked by β-1,4-glycosidic bonds. Cellulose is a polysaccharide (see CARBOHYDRATES), and its glycosidic linkages are cleaved during hydrolysis. If degradation is extensive, cellobiose (the dimer) or glucose is produced. The structure of cellulose is presented in its preferred conformation in Figure 6.

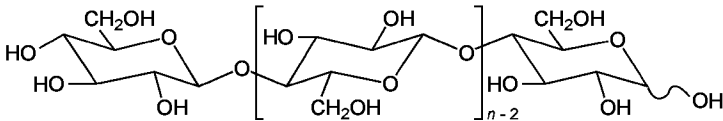


Fig. 6. Chemical structure of cellulose.

The molecular cellulose chains have varying lengths. Measurements of the chain length require a solution of the cotton. Solvents for this purpose include cuprammonium hydroxide solution, phosphoric acid [7664-38-2], nitric acid [7697-37-2], quaternary ammonium bases, cadmium ethylenediamine hydroxide [14874-24-9], cupriethylenediamine hydroxide [111274-71-6] (24), and the current favorite N,N-dimethylacetamide [127-19-5]-lithium chloride [7447-41-8] (DMAC-LiCl). The last, when used in conjunction with gel permeation chromatography (gpc) (25,26) provides both the weight (M_w) and number average (M_n) molecular weight of cellulose in a nondegrading solvent without derivatization. Cotton cellulose generally has a higher molecular

weight than wood cellulose. Data for three cotton varieties are presented in Table 2.

Table 2. Molecular Weights of Upland Cotton Varieties ^a

Variety	M_n	M_w
Deltapine Acala 90	130,000	1,640,000
Deltapine 41	134,000	1,580,000
DES 422	613,000	1,210,000

^a As measured by gpc in DMAC–LiCl (26); based on polystyrene standards.

Supramolecular Structure. The chains of cellulose molecules associate with each other by forming intermolecular hydrogen bonds and hydrophobic bonds. They coalesce to form microfibrils also called crystallites. In cotton, the microfibrils can organize into macrofibrils 60 to 300 nm wide. The macrofibrils are organized into fibers. Cotton fibers have a complex, reversing, helical arrangement of macrofibrils. There are several different forms or polymorphs (cellulose I to IV and X with recent subclasses I α and I β (27,28)), depending on the source and treatment. There are both different unit cells and different packing arrangements in the unit cell. Native cotton is cellulose I. It has been proposed that cotton and other commercial cellulose, such as wood and ramie, are mostly cellulose I β (29). This new and evolving work on the as yet unresolved crystalline natures of native celluloses is detailed elsewhere (see CELLULOSE).

Cotton cellulose is converted to cellulose II by repeated mercerization, the swelling of cellulose in strong alkali (eg, 23% NaOH), followed by rinsing and drying. Typical one-step commercial mercerizing of cotton yarn causes only partial conversion. Cellulose III results from treatment of cellulose with liquid ammonia (ammonia mercerization) or amines. Cellulose III can be made from either cellulose I or II. When treated with water, cellulose III can revert to its parent structure. Cellulose IV can be prepared by treating cellulose I, II, or III in glycerol at temperatures around 260°C. Conversion of the crystal form in cotton fibers to cellulose IV can be effected by heat treatment of ethylamine-treated cotton cellulose in either saturated steam or formamide with minimal fiber degradation (30). Like cellulose III, cellulose IV preparations can revert to their parent structures.

Conversion to cellulose II and cellulose III via caustic mercerization and liquid ammonia treatment are commercial textile processes that are discussed later. Figure 7 shows the characteristic diffractograms (CuK α radiation) of native cellulose, cellulose mercerized with sodium hydroxide, and cellulose treated with liquid ammonia.

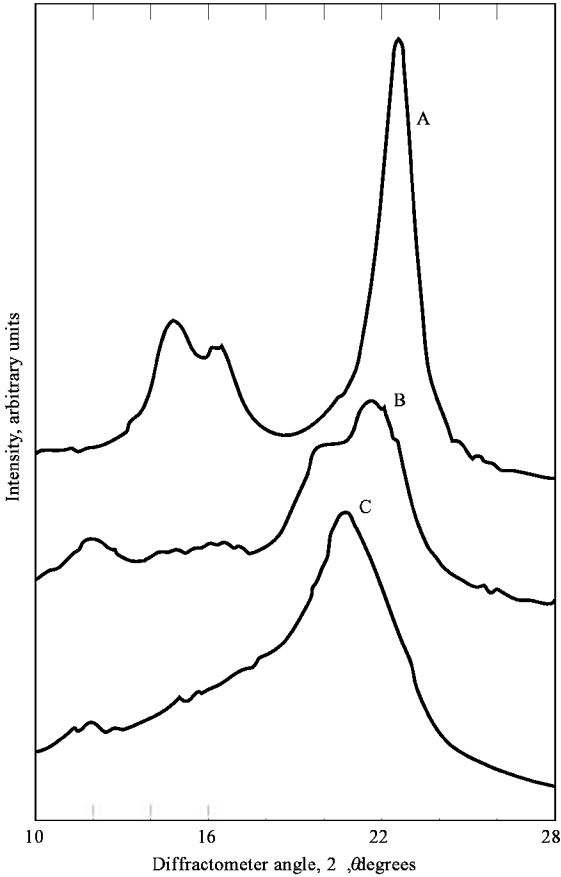


Fig. 7. X-ray diffractograms. A, native; B, NaOH mercerized; C, NH₃ treated.

Pore Structure and Affinity for Water. The cotton fiber is a porous, hydrophilic material that accounts for the comfort of cotton clothing. Moisture is retained tenaciously in cotton. The moisture absorbed from the atmosphere and held under ambient conditions is expressed either as moisture content (amount of moisture as a percentage of original sample mass) or more commonly as moisture regain (amount of moisture as a percentage of oven-dry sample). Under ordinary atmospheric conditions, moisture regain is 7 to 11%. Upon immersion in liquid water the cotton fiber swells and its internal pores fill with water. Pure cotton holds a substantial percentage of its dry weight in water under conditions of centrifugation. Values for the liquid water held depend on the test used. These are approximately 30% for water of imbibition (31) and 50% for the water retention value (32). Centrifugation conditions are less severe in the latter case.

Pores accessible to water molecules are not necessarily accessible to chemical agents. Many uses of cotton, eg, easy care fabric, depend on chemical modification to impart the desired properties. Knowledge of its accessibility under water-swollen conditions to dyes and other chemical agents of various sizes is required for better control of the various chemical treatments applied to cotton textiles. The principle of molecular exclusion via a gpc technique (33) has been used to assess the pore size distribution in cottons after various chemical treatments. Trends for accessibility to sugars of increasing size are depicted in Figure 8. These probes cover the range of molecular sizes of reagents generally used to modify cotton chemically.

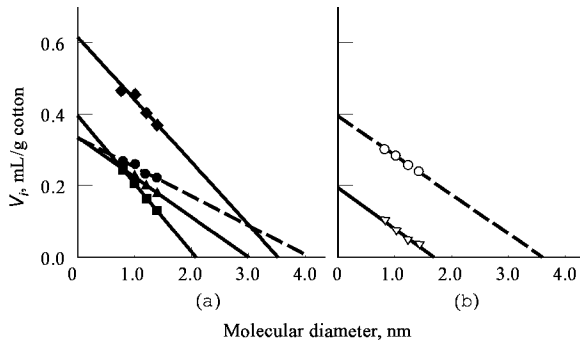


Fig. 8. Internal volume (V_i) that is accessible to sugars as functions of the cotton molecular diameters (33). (a) Batting: ▲, greige; ●, scoured-bleached; ◆, caustic mercerized; ■, liquid ammonia treated. (b) Fabric: ○, scoured-bleached; ▽, cross-linked.

Scouring and bleaching slightly increase the accessible internal volume, liquid ammonia treatment of the scoured-bleached cotton decreases it slightly, caustic mercerization substantially enhances accessibility, and cross-linking to impart durable press properties reduces this accessible internal pore volume substantially.

Availability of Hydroxyl Groups. The chemical structure given in Figure 6 shows the 2-OH, 3-OH, and 6-OH groups that are potential sites for the same chemical reactions that occur with common alcohols. However, the regular occurrences of intermolecular and intramolecular hydrogen bonds in the crystalline regions of cotton cellulose render the involved hydroxyl groups unavailable to chemical agents under mild reaction conditions. Chemical agents that have access to the interior pores of the cotton fiber thus find potential reactive sites unavailable for reaction. Direct information on the availabilities of 2-OH, 3-OH, and 6-OH on accessible surfaces has been obtained from chemical measurements based on the reaction of the cellulose with diethylaminoethyl chloride [2210-36-8] under mild conditions (34,35). The order of decreasing availability of hydroxyl groups in cotton is 2-OH > 6-OH >> 3-OH. Specific values for the relative availability of the 3-OH and 6-OH to the 2-OH depend on the growth (36) and processing (37) history of the fiber. Data on growth and weathering are given in Table 3. Values for relative availability of hydroxyls are maintained throughout the ginning procedure but gradually increase (to ~0.40 and ~0.80 for 3-OH 2-OH and 6-OH 2-OH, respectively) as the fibers are subjected to the stress of processing in the greige mill (37).

Table 3. Relative Availabilities of Hydroxyl Groups of Cotton Cellulose Throughout Growth and Weathering^a

Days postanthesis ^b	Period	3-OH 2-OH	6-OH 2-OH
20	growth: closed boll	0.12	0.49
27		0.05	0.59
34		0.05	0.58
41		0.05	0.59
48		0.06	0.59
62	field weathering: open boll	0.24	0.68
83		0.30	0.75
104		0.31	0.74

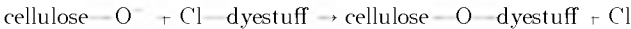
^a Ref. 36.
^b Bolls open shortly after 48 days postanthesis.

Reactions for Practical Objectives.

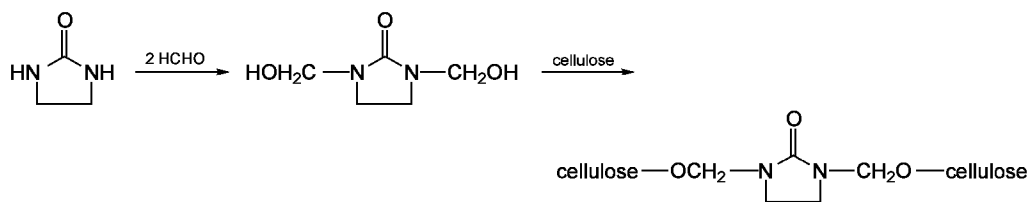
Chemical modification has assisted in building cotton's position in the market-place despite the advent of synthetic fibers.

Mercerization. One of the earliest known modifications of cotton that had commercial potential was mercerization. Traditionally, the process employed a cold concentrated sodium hydroxide treatment of yarn or woven fabric followed by washing and a mild acetic acid neutralization. Maintaining the fabric under tension during the entire procedure was integral to the expected properties. The resultant mercerized cotton of commerce has improved luster and dyeability and, to a lesser extent, improved strength. A variation of this procedure substitutes hot sodium hydroxide that is allowed to cool while the cotton remains immersed in the caustic solution. More thorough initial penetration increases the efficiency of the mercerizing process. If the cotton is allowed to shrink freely during contact with mercerizing caustic, slack mercerization takes place; this technique produces a product with greatly increased stretch (stretch cotton) that has found application in both medical and apparel fields. Effects similar to those from sodium hydroxide mercerization have been produced by exposure of the cotton to volatile primary amines or to ammonia. A procedure that uses liquid ammonia has found commercial adaptation (38). Improvements in luster and strength are similar to those from sodium hydroxide mercerization, but dyeability is not enhanced. A distinct structural difference exists chemically between cotton mercerized in sodium hydroxide and that treated in liquid ammonia, particularly after nonaqueous quenching (39). With both treatments, there is increased accessibility, but differences in dye receptivity presumably result from differences in swelling loci between sodium hydroxide and liquid ammonia treatments.

Etherification. The accessible, available hydroxyl groups on the 2, 3, and 6 positions of the anhydroglucose residue are quite reactive (40) and provide sites for much of the current modification of cotton cellulose to impart special or value-added properties. The two most common classes into which modifications fall include etherification and esterification of the cotton cellulose hydroxyls as well as addition reactions with certain unsaturated compounds to produce cellulose ethers (see CELLULOSE, ETHERS). One large class of cellulose-reactive dyestuffs in commercial use attaches to the cellulose through an alkali-catalyzed etherification by nucleophilic attack of the chlorotriazine moiety of the dyestuff:



Cross-Linking. By far the most important commercial modifications of cotton cellulose remain those that occur through etherification. For example, commercial modification of cotton to impart durable-press, smooth drying, or shrinkage resistance properties involves cross-linking adjacent cellulose chains through amidomethyl ether linkages. This cross-linking is commonly achieved by immersing the fabric in a solution of the agent and an appropriate catalyst, removing the excess liquid by passing the fabric through squeeze rolls (pad), drying in an oven to remove the remaining water, and heating to a high temperature to effect covalent bond formation to cellulose. Such a sequence is called a pad-bake process. Although methylene, or oligomeric, cross-links from a pad-bake formaldehyde treatment have been produced, the resultant fabric exhibits severe strength loss. There is, however, a process for cross-linking cotton-containing garments with formaldehyde in the vapor phase that has found commercial acceptance in the uniform-rental garment market. Most reagents for cross-linking cotton cellulose are difunctional or polyfunctional amidomethylol compounds or amido compounds that have pendent hydroxyls on carbons alpha to the amido nitrogen (see AMINO RESINS). The methol compound is generated by reaction of the amido compound with formaldehyde [50-00-0]. For example, ethyleneurea [120-93-4] reacts as follows (41).



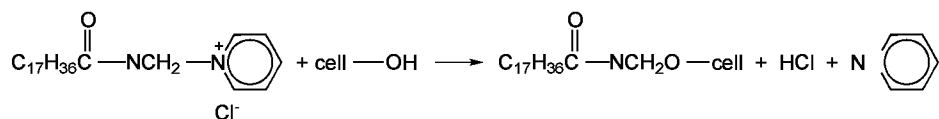
Commercially available cross-linking agents include dimethylolurea [140-95-4], dimethylolethyleneurea [136-84-5], dimethyloldihydroxyethyleneurea [1854-26-8], dimethylolpropyleneurea [3270-74-4], dimethylolalkyl carbamate, tetramethylolacetylenediurea [5395-50-6], methylolated melamine, dimethylolalkyltriazone, dimethoxymethyluron [7327-69-7], dihydroxydimethylethyleneurea [3923-79-3] (dimethylurea–glyoxal adduct), and ethyleneurea–glyoxal [107-22-2] adducts. The cross-linking proceeds via either Lewis or Brønsted catalysis (42) by a carbocation mechanism. Gross effects of the cross-linking are increased resiliency (manifested in wrinkle resistance, smooth drying properties, dimensional stability, and greater shape-holding properties, all of which are important to cotton textiles) and, conversely, reduced extensibility, strength, and moisture regain. These effects are observed with one cross-link per 20 to 25 anhydroglucose residues (43). Liquid ammonia treatment of cotton fabric, followed by cross-linking, attenuates the strength loss as well as an accompanying loss in abrasion resistance. This combination contributed to a reappearance of all-cotton fabrics in the woven shirting sheeting market in the 1970s (44).

Increasing Resiliency. Base-catalyzed reactions of cotton cellulose with either monoepoxides or diepoxides to form cellulose ethers also result in fabrics with increased resiliency. Monoepoxides, believed to result only in cellulose hydroxyalkyl ethers or linear graft polymers (45), produce marked improvement in resiliency under wet conditions, but little improvement under ambient conditions. Difunctional epoxides, capable of cross-linking the cellulose, can produce increases in resiliency under both wet and ambient conditions (46). Besides imparting resiliency through epoxide etherification, oil and water repellency can be imparted by reactions of monomeric perfluoro epoxides with cotton. An epoxide reaction with cotton also allows cotton to accept dyes not traditionally used for cotton. Etherification of cotton with ethyleneimine [2734-98-8] has provided the means for imparting special properties to cotton; the end product depends on the attached group. Another base-catalyzed etherification is the cross-linking reaction between bis(hydroxyethyl) sulfone and cellulose; fabrics possessing increased resiliency under both wet and ambient conditions are obtained. The earliest application of sulfone cross-links to cotton textiles was the reaction of divinyl sulfone under alkaline conditions (47). However, the hazard of working with the vinyl compound led to modifications of the sulfone agent to replace the vinyl groups with more stable precursors such as the β -thiosulfatoethyl, β -sulfatoethyl (48), or β -hydroxyethyl groups (49). Etherification of cotton by divinyl sulfone [77-77-0] and its precursors also forms the basis for a large class of fiber-reactive dyestuffs with the general formula of dyestuff-SO₂CH=CH₂ (50) (see DYES, REACTIVE).

Other Cellulose Ethers. Other cotton cellulose ethers include carboxymethyl, carboxyethyl, hydroxyethyl, carbamoylethyl, cyanoethyl, sulfoethyl, and aminoethyl (aminized cotton) products. Most, with the exception of cyanoethylated and aminized cotton, are of interest in applications requiring solubility or swellability in water or alkali (51). In addition, ethers with pendent acid or basic groups have ion-exchange properties (52). Aminized cotton is of interest because it introduces into the cotton basic groups that provide sites for attachment of acid dyes. Simultaneous aminization of cotton and dyeing with an acid dyestuff marked the first successful attempt at dye attachment to cellulose through an ether linkage (53).

Flameproofing. The flameproofing of cotton is discussed elsewhere (see FLAME RETARDANTS FOR TEXTILES). Although certain cellulose esters such as the ammonium salt of phosphorylated cotton and cellulose phosphate [9015-14-9] are flame resistant, the attachment of most currently used durable polymeric flame retardants for cotton is through an ether linkage to the cellulose at a relatively low degree of substitution (DS). Nondurable flame retardants based on liquid or vapor-phase applications of boric acid [10043-35-3] or methyl borate [121-43-7] are used in treatment of cotton batting for upholstery, bedding, and automotive cushions (54).

Water Repellency. The development of water-repellent cellulose ethers has been reviewed (55) (see WATERPROOFING). A typical example of a commercial etherification for waterproofing cotton is with stearamidomethylpyridinium chloride:



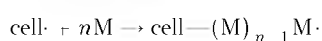
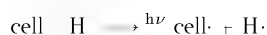
N-substituted, long-chain alkyl monomethylol cyclic ureas have also been used to waterproof cotton through etherification. Other water repellent finishes for cotton are produced by cross-linked silicone films (56). In addition to the polymerization of the phosphorus-containing polymers on cotton to impart flame retardancy and of silicone to impart water repellency, polyfluorinated polymers have been successfully applied to cotton to impart oil repellency. Chemical attachment to the cotton is not necessary for durability; oil repellency occurs because of the low surface energy of the fluorinated surface (57).

Cyanoethylation. One of the earliest examples of etherification of cellulose by an unsaturated compound through vinyl addition is the cyanoethylation of cotton (58). This base-catalyzed reaction with acrylonitrile [107-13-1], a Michael addition, proceeds as follows:



For most textile uses, a DS less than 1 is desirable. Cyanoethylation can impart a wide variety of properties to the cotton fabric such as rot resistance, heat and acid resistance, and receptivity to acid and acetate dyes. Acrylonitrile (qv) has also been radiation-polymerized onto cotton with a ⁶⁰Co source. Microscopical examination of ultrathin sections of the product shows the location of the polymer is within the fiber (59). Examination of the ir spectrum of cotton containing polymerized acrylonitrile indicates grafting does occur at the hydroxyl site of the cellulose (60). Another monomer grafted onto cellulose by irradiation is styrene (qv). Chemical properties, mechanisms, and textile properties of these grafted polymers of cellulose have been summarized (61). Graft polymerization onto cotton has also been induced by both chemical (62) and photochemical (63) initiation (see RADIATION CURING).

Irradiation. The effects of high energy radiation on cotton properties have also been investigated, eg, the effects of gamma radiation on cotton have been described (64–66). Depolymerization of the cellulose occurs with increasing energy absorption; carbonyl formation, carboxyl formation, and chain cleavage occur in the ratio of 20:1:1. With these chemical changes, there is a corresponding increase in solubility in water and alkali and a decrease in fiber strength. The gamma-irradiated cotton has base ion-exchange properties. Irradiation of cotton with near ultraviolet light (325 to 400 nm) causes formation of cellulose free radicals and mild oxidative degradation of the cotton (67). Carbonyl and carboxyl contents of the cotton cellulose increase, and DP and tensile strength decrease, with increasing time of irradiation (68) (see PHOTOCHEMICAL TECHNOLOGY). The induction of cellulose free radicals by near uv irradiation forms the basis for photofinishing with vinyl monomers to produce graft polymers on the cotton:



Another useful reaction of cotton cellulose occurs with an ionized atmosphere. This is essentially a surface reaction. Glow discharge treatment of cotton yarn in air increases water absorbency and strength (69), and surface-dependent properties of cotton fabric are drastically changed by exposure to low temperature–low pressure plasma generated by radio-frequency radiation (70). Because only a few extremely high energy electrons (10 to 15 eV) are generated, ambient temperature is maintained in the chamber. Light microscopy indicates a smoother surface, although scanning electron microscopy shows no change from native cotton. Spectral changes show some oxidation of the cotton, a decreased carbon-to-oxygen ratio. Free radicals similar to those from ⁶⁰Co radiation are formed. In addition, highly charged species are also formed, allowing such usually inert monomers as benzene to be polymerized onto the cotton; there is great capacity for bond cleavage. Plasma treatment gives an increased rate of wetting and drying and produces a

highly absorbent cotton. The cohesiveness and fiber friction of cotton sliver was increased temporarily through air-trace chlorine corona treatments at 95°C and atmospheric pressure (71,72). With a 15-kV electrode voltage at a frequency of 2070 Hz, no chemical effects on the cotton could be noted. Dyeability, hand, and wettability were unaffected. The increase in cohesiveness allows the production of yarns with increased strength, abrasion resistance, and greater spinnability (73). Thus yarns of significantly lower twist can be produced with strength equal to, or higher than, untreated cotton yarns of higher twist.

Insolubilization. Insolubilization of compounds within textiles parallels the history of humanity; the direct dyeing technique for cotton was highly advanced in the Bronze Age (see DYES, NATURAL). With the exception of fiber-reactive dyestuffs discussed earlier, all other cotton dyes, ie, substantive, vat and sulfur, are insolubilized within the fiber, the latter two after an oxidizing step (see DYES AND DYE INTERMEDIATES). Insoluble metal oxides have been used to flameproof cotton. Certain zirconium compounds have been insolubilized on cotton to render the fabric microbial resistant (74) or mildew resistant (75) via a mineral dyeing process (see TEXTILES).

Insolubilization and five other methods for imparting antimicrobial properties to cotton have been described (76). These methods can all be classified under one or more of the chemical reactions of cotton cited earlier; they include fiber reactions to form metastable bonds, grafting through thermosetting agents, formation of coordination compounds, ion-exchange methods, polymer formation with possible grafting, and a regeneration process. A commercialized process for antibacterial cotton fabrics uses peroxide complexes of zirconyl acetate (77).

When exposed to heat, cotton fabrics, like most substances, increase in temperature to an extent that is proportional to their specific heats. Altering the chemical composition of the fabrics such that large amounts of heat are absorbed and released in repeatable cycles of controllable temperature ranges produces fabrics that are described as temperature adaptable. The process insolubilizes poly(ethylene glycol)s cross-linked with methylolamides in the cotton fabric (78). As with flame-retardant cellulose, attachment is through an ether linkage to the cellulose at a relatively low DS.

Esterification. There are both inorganic and organic esters of cellulose (79) (see CELLULOSE ESTERS). Of the three most common inorganic esters, cellulose nitrate [9004-70-0], phosphate, and sulfate, only the cellulose sulfate [9088-06-6] is soluble in water. Cellulose sulfate attains water solubility at a DS of 3, indicating esterification of all three hydroxyls, whereas the sodium salt of cellulose sulfate is soluble in hot and cold water with a DS of only 0.33. Sodium cellulose sulfate is used in applications requiring suspension, thickening, stabilizing, and film-forming properties. The class of phosphonic acid and phosphoric acid dyestuffs attach to cotton through esterification by the phosphonic acid or phosphoric acid group of the dyestuff. Until recently, organic esters of cotton cellulose, with two notable exceptions, were only of academic interest, although partial esterification of cotton by fatty acids has been reported to increase resiliency (80).

Acetylation of cotton to an acetyl content slightly greater than 21° produces a material with greatly increased resistance to fungal and microbiological degradation, in addition to a tolerance of high temperatures not exhibited by native cotton; fibrous appearance and physical properties are unchanged by the acetylation. X-ray diffractograms indicate that, at this extent of substitution, only accessible (noncrystalline) regions of the cotton are involved in the acetylation (81).

In the 1960s, esterification of cotton cellulose by polycarboxylic acids to produce smooth-drying fabrics was investigated (82,83). Catalysis was by partial neutralization of the acid groups. Although improvements in resiliency were obtained, the levels were not commercially acceptable. In the late 1980s, better catalysis was discovered for the ester cross-linking of cellulose; inorganic salts of phosphorus-containing acids were found to give ester cross-links that are durable to multiple home launderings. Because of the improved catalysis, certain tricarboxylic and tetracarboxylic acids have shown promise for commercialization. These acids include 1,2,3,4-butanetetracarboxylic acid [1703-58-8], tricarballic acid [99-14-9], and citric acid [77-92-9] (84). An anhydride mechanism has been proposed for the esterification cross-linking of cellulose. An advantage of the polycarboxylic acids in finishing for durable press is that these agents do not contain formaldehyde and thus do not release formaldehyde during processing or end use. Finishes from polycarboxylic acids are superior to those from other nonformaldehyde agents mentioned earlier, such as epoxides, sulfones, acetals, and cyclic urea derivatives, because they are innocuous and are durable to home laundering.

Economic Aspects

Marketing. There are several routes by which cotton fiber in the United States changes ownership from the grower to its final destination at the spinning mill. The grower may sell cotton directly to a spinning mill under a grower contract or the grower may sell cotton to a gin, broker, commission firm, or shipper. Some growers, after ginning their cotton, may sell through a cooperative organization or may place the cotton in a depository as collateral under the Commodity Credit Corporation Loan Program, to be either withdrawn on repayment of the loan plus interest and storage or forfeited for sale by the government. These intermediate buyers then sell the cotton to foreign or domestic mills that have not purchased the cotton directly from the grower.

World Production and Prices. World production, consumption, and prices are shown in Tables 4 and 5. Through the Universal Cotton Standards of Agreement (86), the official cotton standards of the United States for Upland cotton are recognized by 14 cotton associations and exchanges in 10 major consuming countries of Europe and Asia. The major differences between the U.S. system and those of other countries include the number of grades established and the factors that influence quality, such as variety, moisture, and geographic background. In some instances, staple length is included as an attribute of grade, whereas in the United States, grade and staple length are separate criteria of the classification system.

Table 4. World Production of Cotton^a, Bales^b

Area	1980 ^c	1985 ^c	1990 ^c
Americas	18,784	20,918	23,421
Africa	5,286	5,635	5,741
Europe	13,122	13,890	13,503
Asia–Oceania	26,524	39,421	44,270
Total	63,716	179,863	86,935

^a Ref. 85.

^b 1 bale = 217.7 kg.

^c A year begins August 1 of the year given and ends July 31 of following year.

Table 5. World Consumption and Price of Cotton^a

Factor	1980 ^b	1985 ^b	1990 ^b
consumption, 10 ³ bales ^c	65,134	76,070	86,377
price ^d \$ kg	2.07	1.08	1.82

^a Ref. 85.

^b A year begins August 1 of the year given and ends July 31 of following year.

^c 1 bale = 217.7 kg.

^d "A" index price as reported by Cotlook Ltd.

Health and Safety Factors

Byssinosis. Byssinosis is a controversial form of occupational lung disease that can affect a small number of textile workers from repeated inhalation of the dust generated during the processing of cotton and some other vegetable fibers (eg, flax and soft hemp) (87,88). Views vary radically between those who believe byssinosis to be a fully reversible disease of the airways not associated with any disability to those who believe it may cause progressive and disabling airway narrowing. Cotton dust, an airborne particulate matter released into the atmosphere as cotton is handled or processed in textile processing, is a complex mixture of botanical trash, soil, and microbiological material (ie, bacteria and fungi) (89). The etiological agent and pathogenesis of byssinosis are not known (88,90,91). However, control studies in experimental cardrooms suggest that endotoxin from Gram-negative bacteria is associated to some degree with worker reaction to dust (92). Appropriate engineering controls in cotton textile processing areas for the most part can eliminate incidence of workers' reaction to cotton dust. The U.S. Occupational Safety and Health Administration (OSHA) issued revised standards for occupational exposure to cotton dust in 1985 (93).

Formaldehyde. Formaldehyde is a component of resins used to impart durable-press and other properties to cotton fabrics. It is classified as a "probable human carcinogen," because it has been shown to be an animal carcinogen and there is limited evidence to indicate that it is a carcinogen in humans (94). Sensory irritation of the mucous membranes of the eyes and the respiratory tract, and cellular changes in the nasal cavity are the principal noncancer effects of exposure to low airborne concentrations of formaldehyde (94). Exposure to formaldehyde from cotton textiles is controlled by the chemical technology on low emitting formaldehyde resin technology and nonformaldehyde finishes discussed earlier and by increased ventilation. OSHA has issued standards for control of occupational exposure to formaldehyde (95). The U.S. Consumer Product Safety Commission (CPSC) does not regulate formaldehyde in textiles because its studies did not indicate that there is an acute or chronic health problem due to formaldehyde exposures from textiles.

BIBLIOGRAPHY

"Cotton" in *ECT* 1st ed., Vol. 4, pp. 563–578, by Kyle Ward, Jr., R. B. Evans, Mary L. Rollins, B. Meadows, and I. de Gruy, Southern Regional Research Laboratory, U.S. Department of Agriculture; in *ECT* 2nd ed., Vol. 6, pp. 376–412, by E. Lord, The Cotton Silk & Man-Made Fibres Research Association, Shirley Institute, Manchester; in *ECT* 3rd ed., Vol. 7, pp. 176–195, by B. A. Kottes Andrews and Ines V. De Gruy, Southern Regional Research Center, United States Department of Agriculture.

1. W. L. Balls, *The Development and Properties of Raw Cotton*, A. & C. Black, Ltd., London, 1915.
2. C. B. Purves in E. Ott, H. M. Spurlin, and M. W. Graefflin, eds., *Cellulose and Cellulose Derivatives*, 2nd ed., Pt. 1, Interscience Publishers, Inc., New York, 1954, pp. 29–53.
3. R. D. Preston, *The Molecular Architecture of Plant Cell Walls*, John Wiley & Sons, Inc., New York, 1952.
4. H. B. Brown and J. O. Ware, *Cotton*, 3rd ed., McGraw-Hill Book Co., Inc., New York, 1958.
5. USDA, *World Cotton Situation*, FAS, Washington, D.C., August 1991.
6. *Cotton Varieties Planted, 1991 Crop*, U.S. Agricultural Marketing Service, Memphis, Tenn., August 1991.
7. G. W. Wallingford, *paper presented at the Southern Soil Fertility Conference*, Memphis, Tenn., Oct. 15–18, 1991.
8. G. A. Slater, *Cotton Int. (Memphis)* **43**, 90, 130, 138 (1976).
9. J. P. Dean, *Agric. Res.* **24**(12), 8 (June 1976).
10. *Cotton Int. (Memphis)* **43**, 55, 58 (1976).
11. U.S. Bureau of Census, *Cotton Production in the United States; Crop of 1970*, U.S. Government Printing Office, Washington, D.C., 1971.
12. E. H. Glade and M. D. Johnson, *Cotton Ginning Charges, Harvesting Practices, and Selected Marketing Costs, 1990/1991 Season*, Staff Report **90–50**, U.S. Department of Agriculture, Economic Research Service, Washington, D.C., 1991.
13. W. S. Anthony, *Cotton Gin Oil Mill Press* **91**(20), 8 (1990).
14. "Standard Test Method for Length and Length Distribution of Cotton Fibers (Array Method)," ASTM D1440-90, *Annual Book of ASTM Standards*, Vol. 07.01., ASTM, Philadelphia, Pa., 1991, pp. 390–395.
15. *U.S. Cotton Handbook*, Cotton Council International and National Cotton Council of America, Washington, D.C., 1976.
16. "Standard Test Method for Breaking Strength and Elongation of Cotton Fibers (Flat Bundle Method)," ASTM D1445-90, *Annual Book of ASTM Standards*, Vol. 07.01, ASTM, Philadelphia, Pa., 1991, pp. 405–409.
17. *Textile World Manmade Fiber Chart, Textile World*, McGraw-Hill Book Co., Inc., New York, Aug. 1988.
18. D. P. Thibodeaux and J. B. Price, *Melliand Textilber* **70**, 243 (1989).
19. E. Lord, *The Origin and Assessment of Cotton Fiber Maturity*, 2nd ed., The International Institute for Cotton, Manchester, UK, 1982.
20. R. J. Kohel and C. F. Lewis, eds., *Cotton*, Agronomy No. **24**, American Society of Agronomy, Crop Science Society of America, Inc., Soil Science Society of America, Inc., Madison, Wis., 1984, pp. 437–465.
21. M. L. Rollins, *Forest Products J.* **18**(2), 91 (1968).
22. J. O. Warwicker, R. Jeffries, R. L. Colbran, and R. N. Robinson, *A Review of the Literature on the Effect of Caustic Soda and Other Swelling Agents on the Fine Structure of Cotton*, Shirley Institute Pamphlet No. 93, Didsbury, Manchester, UK, 1966.
23. M. L. Rollins in D. S. Hamby, ed., *The Cotton Fiber, American Cotton Handbook*, 3rd ed., Interscience Publishers, New York, 1965.
24. D. C. Johnson in T. P. Nevel and S. H. Zeronian, eds., *Cellulose Chemistry and Its Applications*, Ellis Horwood, Chichester, UK, 1985, pp. 181–201.
25. J. D. Timpa, *J. Ag. Food Chem.* **39**(2), 270 (1991).
26. J. D. Timpa and H. H. Ramey, *Textile Res. J.* **59**, 611 (1989).
27. J. Sugiyama, T. Okano, H. Yamamoto, and F. Horii, *Macromolecules* **23**, 3196 (1990).
28. J. Sugiyama, R. Vuong, and H. Chanzy, *Macromolecules* **24**, 4168 (1991).
29. R. H. Atalla and D. L. VanderHart, *Science* **223**, 1465 (1984).
30. S. H. Zeronian and H.-S. Ryu, *J. Appl. Polym. Sci.* **33**, 2587 (1987).
31. H. M. Welo, H. M. Ziffle, and A. W. McDonald, *Textile Res. J.* **22**, 261 (1952).
32. G. Jayme, *Tappi* **41**, 180A (1958).
33. N. R. Bertoniere in J. F. Kennedy, G. O. Phillips, and P. A. Williams, eds., *Cellulose, Structural and Functional Aspects*, Ellis Horwood Series in Polymer Science and Technology, Chichester, UK, 1989, pp. 99–104.
34. S. P. Rowland and N. R. Bertoniere in Ref. 24, pp. 112–137.
35. N. R. Bertoniere and S. H. Zeronian in R. H. Attala, ed., in *The Structures of Cellulose, ACS Symposium Series 340*, Washington, D.C., 1987, pp. 255–271.
36. S. P. Rowland and P. S. Howley, *J. Polym. Sci. Polym. Chem. Ed.* **23**, 183 (1985).
37. N. R. Bertoniere, P. S. Howley, S. E. Hughs, W. S. Anthony, and G. F. Ruppenicker, *Proceedings of the Beltwide Cotton Conference*, Vol. 3, Nashville, Tenn., 1992, p. 1267.
38. **Brit. Pat. 1,084,612** (September 27, 1967) (to Sentralinstitut for Industriell Forskning and Norsk Tekstilforskningsinstitut); **Brit. Pat. 1,136,417** (December 11, 1968), R. M. Gailley (to J. and P. Coats, Ltd.).
39. T. A. Calamari, Jr., S. P. Schreiber, A. S. Cooper, and W. A. Reeves, *Text. Chem. Color.* **3**, 235 (1971).
40. E. Heuser, *Text. Res. J.* **20**, 828 (1950); T. E. Timell, *Sven Papperstidn.* **56**, 483 (1953).
41. J. G. Frick, Jr., *Chem. Technol.* **1**, 100 (1971).

42. A. G. Pierce, Jr., R. M. Reinhardt, and R. M. H. Kullman, *Text. Res. J.* **46**, 420 (1976).

43. J. G. Frick, Jr., B. A. Kottes Andrews, and J. D. Reid, *Text. Res. J.* **30**, 495 (1960).

44. S. A. Heap, *Colourage* **24**(7), 15 (1977).

45. J. B. McKelvey, B. G. Weber, and E. Klein, *Text. Res. J.* **29**, 918 (1959).

46. R. R. Benerito, R. J. Berni, J. B. McKelvey, and B. G. Burgess, *J. Appl. Polym. Sci. Part A* **1**, 3407 (1963).

47. **U.S. Pat. 2,524,399** (October 3, 1950), D. L. Schoene and V. S. Chambers (to U.S. Rubber Co.).

48. G. C. Tesoro, P. Linden, and S. B. Sello, *Text. Res. J.* **31**, 283 (1961).

49. **Can. Pat. 625,790** (August 15, 1961), R. O. Steele (to Rohm and Haas Co.).

50. R. H. Peters, *Textile Chemistry III, The Physical Chemistry of Dyeing*, Elsevier Scientific Publishing Co., Inc., New York, 1975, pp. 624–629.

51. L. A. Graham and C. A. Suratt, *Am. Dyest. Rep.* **67**(7), 36 (1978).

52. J. D. Guthrie, *Ind. Eng. Chem.* **44**, 2187 (1952).

53. J. D. Guthrie, *Am. Dyest. Rep.* **41**(1), P13, 30 (1952).

54. N. B. Knoepfler in V. M. Bhatnagar, ed., *Proceedings of the European Conference on Flammability and Fire Retardants*, Technomic Publishing Co., Westport, Conn., 1980, p. 115.

55. J. T. Marsh, *An Introduction to Textile Finishing*, 2nd ed., Chapman and Hall, London, 1966, pp. 458–494.

56. C. M. Welch, J. B. Bullock, and M. F. Margavio, *Text. Res. J.* **37**, 324 (1967).

57. E. J. Grajeck and W. H. Petersen, *Text. Res. J.* **32**, 320 (1962).

58. G. C. Daul, R. M. Reinhardt, and J. D. Reid, *Text. Res. J.* **25**, 246 (1955).

59. J. C. Arthur, Jr., and R. J. Demint, *Text. Res. J.* **30**, 505 (1960).

60. J. C. Arthur, Jr., and R. J. Demint, *Text. Res. J.* **31**, 988 (1961).

61. J. C. Arthur, Jr., and F. A. Blouin, *Am. Dyest. Rep.* **51** (26), 1024 (1962).

62. H. H. St. Mard, C. Hamalainen, and A. S. Cooper, Jr., *Am. Dyest. Rep.* **56**(5), 24 (1967).

63. A. H. Reine, N. A. Portnoy, and J. C. Arthur, Jr., *Text. Res. J.* **43**, 638 (1973).

64. F. A. Blouin and J. C. Arthur, Jr., *Text. Res. J.* **28**, 198 (1958).

65. J. C. Arthur, Jr., *Text. Res. J.* **28**, 204 (1958).

66. R. J. Demint and J. C. Arthur, Jr., *Text. Res. J.* **29**, 276 (1959).

67. J. C. Arthur, Jr., *Proceedings of the 13th Cotton Utilization Conference*, ARS-S-64, U.S. Agricultural Research Service, Southern Research Center, New Orleans, La., 1975, p. 6.

68. J. C. Arthur, Jr., and O. Hinojosa, *Appl. Polym. Symp.* **26**, 147 (1975).

69. R. B. Stone and J. R. Barrett, Jr., *Text. Bull.* **88**(1), 66 (1962).

70. H. Z. Jung, T. L. Ward, and R. R. Benerito, *Text. Res. J.* **47**, 217 (1977).

71. W. J. Thorsen, *Text. Res. J.* **41**, 331 (1971).

72. *Ibid.*, p. 455.

73. D. P. Thibodeaux and H. R. Copeland, 15th Textile Chemical Processing Conference, USDA Southern Regional Research Center, New Orleans, La., 1975, pp. 36–40.

74. C. J. Conner, A. S. Cooper, W. A. Reeves, and B. J. Trask, *Text. Res. J.* **34**, 347 (1964).

75. C. J. Conner, G. S. Danna, A. S. Cooper, and W. A. Reeves, *Text. Res. J.* **37**, 94 (1967).

76. D. D. Gagliardi, *Am. Dyest. Rep.* **51**, P49 (1962).

77. T. L. Vigo, G. F. Danna, and C. M. Welch, *Text. Chem. Color.* **9**, 77 (1977).

78. **U.S. Pat. 4,851,291**, (July 25, 1989), T. L. Vigo, C. M. Frost, J. S. Bruno, and G. F. Danna (to U.S. Department of Agriculture).

79. R. M. Reinhardt and J. D. Reid, *Text. Res. J.* **27**, 59 (1957).

80. J. B. McKelvey, R. J. Berni, and R. R. Benerito, *Text. Res. J.* **34**, 1102 (1964).

81. C. F. Goldthwait, E. M. Buras, Jr., and A. S. Cooper, Jr., *Text. Res. J.* **21**, 831 (1951).

82. D. D. Gagliardi and F. B. Shippee, *Am Dyest. Rep.* **52**, P300 (1963).

83. **U.S. Pat. 3,526,048** (September 1, 1970), S. P. Rowland, C. M. Welch, and M. A. F. Brannan (to U.S. Department of Agriculture).

84. **U.S. Pat. 4,820,307** (April 11, 1989), C. M. Welch and B. K. Andrews (to U.S. Department of Agriculture).

85. International Cotton Advisory Committee, *Cotton-World Statistics*, April 1988 and October 1991.

86. *The Classification of Cotton*, Miscellaneous Publication No. **310**, U.S. Department of Agriculture, Washington, D.C., January 1965, pp. 11–15.

87. R. Rylander and co-workers, *Br. J. Ind. Med.* **44**, 577 (1987).

88. C. A. C. Pickering in R. R. Jacobs, P. J. Wakelyn, and L. N. Domelsmith, eds., *Proceedings of the 15th Cotton Dust Research Conference*, National Cotton Council, Memphis, Tenn., 1991, pp. 298–299.

89. P. J. Wakelyn, G. A. Greenblatt, D. F. Brown, and V. W. Tripp, *Am. Ind. Hyg. Assoc. J.* **37**, 22 (1976).

90. M. S. Rohrbach in Ref. 2, pp. 300–306.

91. P. J. Nichols in Ref. 2, pp. 307–320.

92. R. M. Castellan, S. A. Olenchock, K. B. Kingsley, and J. L. Hankinson, *New Engl. J. Med.* **317**, 605 (1987).

93. *Fed. Reg.* **50**, 51120–51179 (Dec. 13, 1985) (29 CFR 1910. 1043).

94. *Formaldehyde Risk Assessment Update*, Office of Toxic Substances, U.S. Environmental Protection Agency, Washington, D.C., June 11, 1991.

95. *Fed. Reg.* **52**, 46168–46312 (Dec. 4, 1987) (29 CFR 1910.1048).

Noelie R. Bertoniere
Southern Regional Research Center, USDA

COTTONSEED.

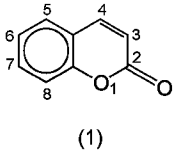
See VEGETABLE OILS.

COUGH REMEDIES.

See EXPECTORANTS, ANTITUSSIVES, AND RELATED AGENTS.

Coumarin

Coumarin [91-64-5], 2*H*-1-benzopyran-2-one; 1,2-Benzopyrone, C₉H O₂ (1), is one of the most important aroma chemicals having unique characteristics not only because of its haylike bittersweet odor, but also because of its quality as a perfume fixative.



Coumarin is widely distributed in the plant kingdom but most of it has been produced synthetically for many years for commercial uses. In addition to its use in the perfumery, cosmetic, and related industries, coumarin has several other industrial applications. Formerly, large quantities of coumarin were used in the food industry mostly associated with vanillin for flavoring chocolates, baked goods, and in the confection of cream soda flavored beverages, but since 1954 its use in food has been suspended in the United States. In a statement of general policy under the Federal Food, Drug, and Cosmetic Act, foods containing coumarin are legally regarded as adulterated (1). Coumarin is the parent substance of a large group of derivatives, many of which occur naturally and some of which are of economic significance.

Occurrence

Coumarin was first isolated by Vogel in 1820 by extraction from tonka beans (*Dipteryx odorata*) (2). It was subsequently identified in a large number of plants belonging to many different families. Its better known occurrences are in sweet clover (*Melilotus officinalis* and *alba*), woodruff (*Asperula odorata*), cassia (*Cinnamorum cassia*), melilot (*Melilotus officinalus*), lavender (*Lavandum officinalus*), and balsam of Peru (*Myroxylon pereirae*) (3,4). Several reviews have been published on the distribution of natural coumarins and their biochemistry (5,6).

Physical Properties

Coumarin is usually sold in the form of colorless shiny leaflets or rhombic crystals. Its ir (7), uv (8), Raman (9), and nmr spectra (10) are known. Physical constants appear in Table 1. Tables 2 and 3 give the solubility of coumarin in various water mixtures and solvents.

Table 1. Physical Constants of Coumarin

Constant	Value	Reference
mol wt	146.14	
mp, °C	71	11,12
bp, °C		
at 100 kPa ^a	301.1	13
at 2.7 kPa ^a	170.4	13
at 0.7 kPa ^a	138.5	13
density at 100 °C, g cm ³	1.178	14

^a To convert kPa to mm Hg, multiply by 7.5.

Table 2. Solubility of Coumarin in Ethanol–Water Mixtures^a

Ethanol, vol ° _o	Solubility, g 100 mL soln. at			
	10°C	20°C	30°C	40°C
0	0.13	0.17	0.29	0.39
25	0.28	0.50	0.97	1.93
50	2.00	3.71	7.56	19.02
70	5.62	10.04	19.3	47.00

^a Ref. 15.

Table 3. Solubility of Coumarin in Various Solvents^a

Solvent	Solubility, g 100 g solvent	Temperature, °C
water	0.25	25
water	2.0	100
chloroform	49.4	25
pyridine	87.7	20–25

^a Ref. 16.

Chemical Properties

The chemical properties of coumarin are those of the lactone of an α,β -unsaturated aromatic acid.

Hydrolysis. The lactone is easily hydrolyzed by alkalies to the corresponding salts of coumarinic acid or *o*-hydroxy-*cis*-cinnamic acid [495-79-4]. Coumarinic acid salts are odorless. Coumarinic acid and salts revert to coumarin upon acidification with inorganic acids. Alkaline fusion of coumarin yields salts of salicylic and acetic acids.

Hydrogenation. Hydrogenation of coumarin gives several different products depending on experimental conditions. Hydrogenation with a Raney nickel catalyst under moderate conditions yields 3,4-dihydrocoumarin [119-84-6] (17) but continued hydrogenation especially at higher temperatures leads to the formation of the saturated octahydrocoumarin [4430-31-3]. Hexahydrochroman [5655-24-3] and polymeric products can also be formed under more severe conditions (18). 3,4-Dihydrocoumarin is also obtained selectively by hydrogenation over a platinum sulfide catalyst (19). Hydrogenation with copper chromite catalyst at high temperature gives 3-(*o*-hydroxyphenyl)-1-propanol [1481-92-1] with a high yield (20).

Reduction. Coumarin is reduced to *o*-hydroxycinnamyl alcohol by reaction with lithium aluminum hydride (21). By reaction with diborane coumarin gives *o*-allylphenol [1745-81-9] (22).

Bisulfite Reaction. Coumarin combines readily with sodium bisulfite solutions to form soluble sodium 3- or 4-hydrosulfonates (23). Coumarin can be regenerated by acidification and this method has been used for its purification.

Halogenation. Coumarin reacts with bromine under moderate conditions to give 3,4-dibromocoumarin [42974-18-5] (24). The 3-bromocoumarin [939-18-4] and 3,6-dibromocoumarin [58309-97-0] are formed under more drastic conditions (25). 3-Chlorocoumarin [9245-5] is formed by reaction with chlorine in dichloroethane (26) or without solvent (27).

Oxidation. Coumarin is not readily oxidized by chromic acid but, by action of the Fenton's reagent, it is converted into 7-hydroxycoumarin (umbelliferone) [93-35-6] (28).

Sulfonation. Fuming sulfuric acid reacts with coumarin to give coumarin-6-sulfonic acid [27279-86-3] at moderate temperature and coumarin-3,6-disulfonic acid [69089-38-9] at higher temperature.

Nitration. Fuming nitric acid forms 6-nitrocoumarin [2725-81-7] (25).

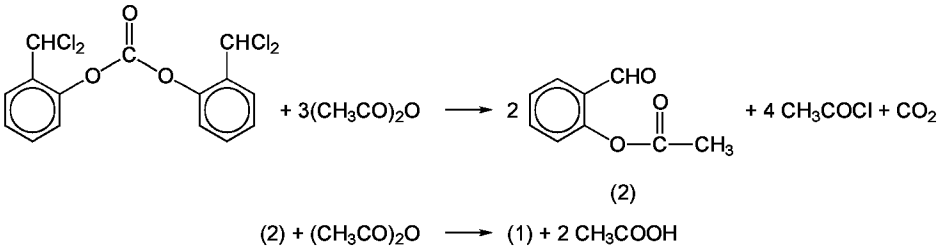
Methylation. Methylating agents such as methyl sulfate and methyl iodide react with coumarin in the presence of sodium hydride to give methyl 2-methoxycinnamate [15854-58-7] (29).

Dimerization. A coumarin dimer is formed by prolonged exposure of coumarin to sunlight or uv radiation. Photodimerization is also catalyzed by boron trifluoride (30).

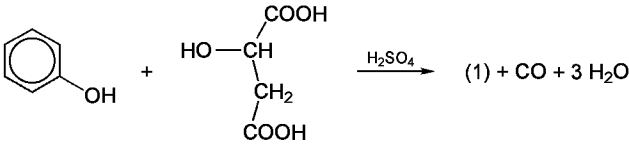
Methods of Preparation

Until the late 1890s, coumarin was obtained commercially from only natural sources by extraction from tonka beans and deer tongue. Then synthetic methods of preparation and industrial manufacturing processes were discovered and developed starting principally from *o*-cresol, phenol, and salicylaldehyde. Various methods can be used to obtain coumarin from each of these starting materials.

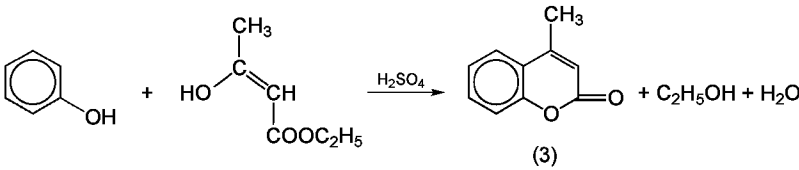
From *o*-Cresol. The Raschig method, starting with *o*-cresol [9548-7], was discovered in 1909 (31). The hydroxyl group of *o*-cresol is first protected by a phosphate or, preferably, a carbonate (32) group and the methyl group is converted into a benzal chloride intermediate by dichlorination. The α,α -dichlorocresyl ester then reacts with an alkali acetate in an alkali fusion reaction (31) or with acetic anhydride in the presence of a metal catalyst such as cobalt oxide (33) to yield *o*-acetylsalicylaldehyde (2), acetyl chloride, and CO₂. Ring closure of *o*-acetylsalicylaldehyde [5663-67-2] with acetic anhydride gives coumarin and acetic acid.



From Phenol. In the type of condensation discovered by von Pechmann in 1883, coumarin is formed by reaction of phenol [108-95-2] with malic (34), maleic, or fumaric acids (35–38) in the presence of concentrated sulfuric acid.



The Pechmann reaction has found extensive applications for the synthesis of numerous coumarin derivatives (39). Coumarin derivatives substituted in the pyrone ring can be obtained by condensing phenol with beta-ketoesters. For example, 4-methylcoumarin (3) is obtained with ethyl acetoacetate [141-97-9].



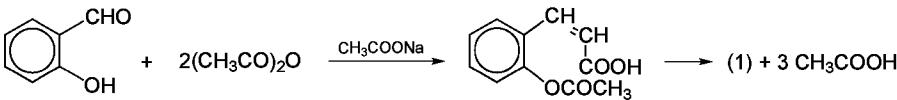
Coumarin can also be formed by the reaction of phenol with diketene (40). Similarly, diphenols can react with hydroxycarboxylic acids or beta-ketoesters to give hydroxycoumarin derivatives. The reaction of resorcinol with malic acid produces umbelliferone (7-hydroxycoumarin) and its reaction with ethyl acetoacetate gives beta-methylumbelliferone (7-hydroxy-4-methylcoumarin).

Phenol reacts with some acrylic acid derivatives to produce coumarin. With 3-ethoxyacrylic acid chloride, phenol gives phenyl ethoxyacrylate that cyclizes into coumarin on treatment with sulfuric acid (41). Coumarin is also formed by reaction of phenol with methyl acrylate in acidic medium and in the presence of air (42).

Hydroxycoumarins can be obtained by reaction of methyl acrylate [96-33-3] with diphenols in the presence of aluminum chloride followed by dehydrogenation with palladium on carbon (43).

From Salicylaldehyde.

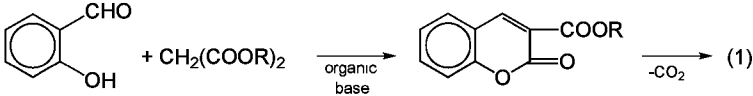
Perkin Reaction. Perkin first synthesized coumarin in 1868 by reaction of the sodium salt of salicylaldehyde with acetic anhydride (44) and it was found later that the reaction could be made from salicylaldehyde [90-02-8] itself by using sodium acetate as a catalyst, through the intermediary of *cis*-*o*-acetoxy-cinnamic acid [50363-92-3].



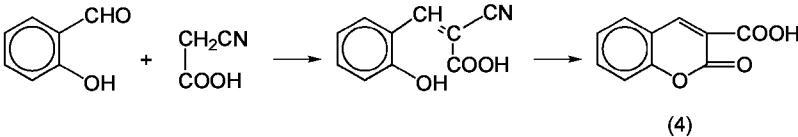
This reaction was also extended to other aromatic aldehydes for the preparation of α,β-unsaturated carboxylic acids. Several mechanisms of the reaction have been proposed (45). The most accepted mechanism involves the reaction of the aldehyde with the enol form of the acid anhydride which is promoted by the presence of the sodium salt or of another base. The resulting reaction product is then dehydrated into an unsaturated carboxylic acid.

The Perkin reaction is of importance for the industrial production of coumarin and a number of modifications have been studied to improve it, such as addition of a trace of iodine (46); addition of oxides or salts of metals such as iron, nickel, manganese, or cobalt (47); addition of catalytic amounts of pyridine (48) or piperidine (49); replacement of sodium acetate by potassium carbonate (50,51) or by cesium acetate (52); and use of alkali metal biacetate (53).

Knoevenagel Reaction. 3-Substituted coumarins can be synthesized by the Knoevenagel reaction (54), which involves the condensation of *o*-hydroxyaldehydes such as salicylaldehyde with acetic acid derivatives containing an active methylene group such as acetoacetic acid, malonic acid [141-82-2], cyanoacetic acid, and their esters. Ammonia or organic bases such as pyridine, piperidine, and primary and secondary amines are used as catalysts (55). Removal of the substituted group in the 3-position by heating or hydrolysis can produce coumarin. Thus coumarin 3-carboxylic acid obtained by the condensation of salicylaldehyde with malonic acid (54) is decarboxylated into coumarin by heating to 290°C. The decarboxylation reaction can be done at a lower temperature and with a better yield in the presence of mercuric salts (56):



where R = H or C₂H₅. The coumarin 3-carboxylic acid [531-81-7] (4) is also obtained by hydrolysis of the cyano group resulting from the condensation of salicylaldehyde with cyanoacetic acid [372-09-8] (57).



Other Methods. Several other methods have been published, including cyclodehydrogenation of oxocyclohexane propionate (58); dehydrogenation of 3,4-dihydrocoumarin obtained by the Baeyer-Villiger oxidation of 1-indanone (59); and the dehydrogenation reaction can be made with sulfur (60) or in the vapor phase with a metal oxide catalyst (61).

Purification and Shipment

In order to be suitable for perfumary uses, synthetic coumarin must be treated to a high degree of purity. Normal purification processes include fractional distillation under high vacuum and crystallization from suitable solvents such as methanol or ethanol. Additional treatments have been proposed prior to distillation including heating with 10–20° concentrated sulfuric acid, neutralizing, and washing with warm water (62); dissolving in 80° concentrated sulfuric acid and treating with an oxidizing gas (63); and treating with aqueous sodium hydroxide under reflux, separating the organic layer, and washing (64).

Coumarin has also been purified by vacuum azeotropic distillation with polyhydroxy alcohols such as triethylene glycol (65). Purification by zone melting techniques has been described in several publications (66–68).

Commercial coumarin is packaged in polyethylene bags and shipped in fiberboard drums. Coumarin is stable in storage and does not require any special handling precautions.

Economic Aspects

Data on the production, sales, and value of synthetic coumarin were published by the U.S. Tariff Commission until 1967 (Table 4). Later data are not available because the number of producers dropped to three or fewer. In 1992, Rhone-Poulenc was the only coumarin producer both in the United States and in Western Europe. Commercial quantities of coumarin have also been produced in China and, to a lesser extent, in the former Soviet Union.

Table 4. U.S. Production and Sales of Synthetic Coumarin

Year	Production, t	Sales		
		Quantity, t	Unit price, \$ /kg	Value, 10 ³ \$
1940	112	99	5.14	508
1946	163	149	5.51	822
1956	354	299	6.68	2000
1967	520	510	4.37	2221
1977	>544	>544	9.92	>5400
1990	>500	>500	15.43	>7700

Specifications

An ir spectrum for identification of coumarin was published by the Scientific Section of the Essential Oil Association of the United States (EOA). Manufacturers' specifications for commercial coumarin are (69)

- Appearance: crystalline powder
- Color of molten product: 70 Hazen maximum
- Odor: characteristic
- Melting point (capillary): 68–70°C
- Assay: 99% minimum

Health and Safety

The acute and subchronic toxicity of coumarin have been well described. The oral LD₅₀ in the rat has been reported to be 293–680 mg/kg (70,71) but no adverse effects have been observed at doses up to 250 ppm for 90 days in the diet of rats; liver damage was noted at 2500 ppm. A subchronic oral toxicity study of coumarin administered was conducted for the National Toxicology Program (NTP) in 1981 (72). The results of this study, conducted for 13 weeks using gavage doses from 19 to 300 mg/kg d, indicated significant decreases in body weight gain and dose-related changes in clinical pathology parameters indicative of hepatic injury at the two highest dose levels, 150 and 300 mg/kg d. Increased liver weight and a histopathologic diagnosis of toxic hepatitis were reported at these two dose levels.

In contrast, the chronic toxicity and potential carcinogenicity of coumarin have been surrounded by scientific debate. Since 1954, coumarin has been classified by the FDA as a toxic substance and its use in foods was banned (1). In a two-year rat study reported in 1967 (73), no effects were observed at 1000 ppm in the diet, but growth retardation and hepatic toxicity (cholangiofibrosis, bile duct proliferation, and focal hepatic necrosis) were reported at 2500 and 5000 ppm. A chronic study in 1973 (74) in which cholangiocarcinomas were reported in rats fed doses of 5000 and 6000 ppm in the diet was inadvertently interpreted as two separate studies during a review of the International Agency for Research on Cancer (IARC) in 1976 (75), leading IARC to classify coumarin as a carcinogen. Further, the interpretation of the histopathologic findings as neoplastic in the 1973 study (74) was questioned in 1979 (76). In 1987, IARC reclassified coumarin as a Group 3 chemical, indicating limited evidence of carcinogenicity in animals and insufficient evidence in humans (77). A more recent chronic study which included an *in utero* exposure phase resulted in hepatotoxicity but no dose-related increase in tumors following two-year administration of 2000 ppm in the diet of rats (78). A two-year gavage study in rats and mice has been conducted for the NTP, although data are not yet available.

Although no quantitative data were presented, rabbits dosed orally and dermally with coumarin showed similar patterns in the urinary excretion of coumarin metabolites (79). Metabolism of coumarin when fed to animals and humans occurs essentially by hydroxylation at all 6-ring positions, although the 7-position appears to be the primary site for many species. In addition, the heterocyclic ring may be opened to yield *o*-hydroxyphenylacetic acid and *o*-phenoxyphenylacetic acid. The 7-hydroxycoumarin is much more prominent in humans (79%) and baboons (60%) than in rats (<1%), rabbits (12%), and dogs (3%) (76,80). The species difference in the metabolism of coumarin may have important toxicologic implications. *In vitro* data demonstrate that the predominant metabolite in rats, *o*-hydroxyphenylacetic acid, inhibits glucose-6-phosphatase in the liver, whereas coumarin and its principal metabolite in humans, 7-hydroxycoumarin, were inactive in this *in vivo* glucose-6-phosphatase inhibition assay (81). Available data from *in vivo* studies suggest that the degree of susceptibility to the hepatotoxic effects of coumarin is strongly correlated with the inability of the exposed species to make 7-hydroxycoumarin, ie, those species with the least ability to hydroxylate at the 7-position have the greatest degree of hepatotoxic response (76). It is therefore possible that humans could be less susceptible than the rat to the hepatotoxic action of coumarin.

Uses

Because of its unique sweet note and stability, coumarin has long been recognized as an important raw material in the fragrance industry (see PERFUMES). It is widely used in hand soaps, detergents, lotions, and perfumes at concentrations usually extending from 0.01 to 0.8%. It is normally associated in perfumery with herbaceous odors and enters in the formulation of fern (Fougere) and Chypre-type fragrances. It is used as an odor enhancer to achieve a long lasting effect when combined with natural essential oils such as lavender, citrus, rosemary, oak moss, etc (see OILS, ESSENTIAL). Coumarin is used in tobacco to enhance its natural aroma. It is also applied in large quantities to give pleasant aromas to household materials and industrial products or to mask unpleasant odors (see ODOR MODIFICATION).

In other fields, coumarin has a significant use in the electroplating industry, mostly in the automotive area, to provide high polished quality to chrome plated steel (see ELECTROPLATING) but this use is presently declining. Coumarin and some of its derivatives have been tested in pharmacology for treatment of schizophrenia (82), or of microcirculation disorders and angiopathic ulcers (83), and also for treatment of high protein edemas in animals (84).

Derivatives

A large number of coumarin derivatives have been identified in plants and many of them have been synthesized and studied for their physiological activity. Only a few are mentioned here because of their economic significance.

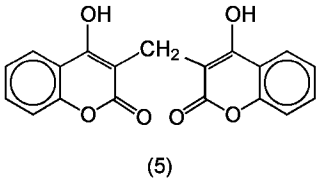
3,4-Dihydrocoumarin [119-84-6] is prepared by catalytic hydrogenation of coumarin. It is also used in the perfumery industry for its haylike odor. It is less powerful than coumarin but its higher solubility in alcohol may make it preferable in some applications. It has GRAS status and can be used as a food flavor ingredient with a sweet caramel-like taste. The U.S. market for dihydrocoumarin is estimated at 50 t/yr at a price of \$34/kg. The U.S. producers are Givaudan and Arsynco.

3-Methylcoumarin [2445-82-1] and 6-methylcoumarin [92-48-8] have some use in the perfume industry. The 6-methyl derivative is permitted in flavor compositions.

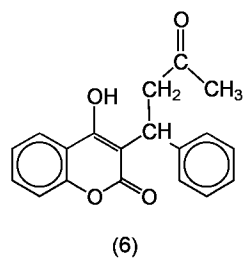
4-Hydroxycoumarin [1076-38-6] can be synthesized by cyclization of acetyl methyl salicylate. It is a coumarin metabolite occurring in spoiled hay. Derivatives of 4-hydroxycoumarin such as dicoumarol [66-76-2], warfarin [81-81-2], cyclocoumarol [518-20-7], ethylbis-coumaracetate [548-00-5], and bis-4-hydroxycoumarin [25892-93-7] are synthetic blood anticoagulants (see BLOOD, COAGULANTS AND ANTICOAGULANTS).

7-Hydroxycoumarin [93-35-6], known as umbelliferone, occurs naturally in gum resins of umbelliferae and is an important coumarin metabolite. It is readily manufactured from resorcinol and maleic or fumaric acid. Umbelliferone and β-methylumbelliferone are used as fluorescent brighteners.

Dicoumarol [66-76-2] (5) was isolated from spoiled sweet clover hay. It is prepared synthetically by reaction of 4-hydroxycoumarin with formaldehyde (85). It is used in anticoagulant therapy often associated with heparin.



Warfarin [81-81-2] (6) is prepared by the Michael condensation of benzylidene acetone with 4-hydroxycoumarin (86). It is used as a rodenticide (see POISONS, COMMERCIAL) and in anticoagulant therapy.



BIBLIOGRAPHY

"Coumarin" in *ECT* 1st ed., Vol. 4, pp. 588–593, by O. DeGarmo, Monsanto Chemical Co.; in *ECT* 2nd ed., Vol. 6, pp. 425–433, by O. DeGarmo and P. Raizman, Monsanto Co.; in *ECT* 3rd ed., Vol. 7, pp. 196–206, by W. C. Meuly, Rhodia Inc.

1. *Fed. Reg.* **19**, 1239 (Mar. 5, 1954); *Code of Federal Regulations*, Title 21, section 189.130, Washington, D.C., 1989.
2. Vogel, *Gilberts Annalen der Physik* **64**, 163 (1820).
3. E. Spath, *Berichte* **70A**, 83 (1937).
4. E. Gildemeister and F. Hoffman, *Die Atherischen Ole*, Vol. IIId. Akademic Verlag, Berlin, 1966, p. 591.
5. S. A. Brown, *Biochem. Plants*, Vol. 7, Academic Press, Inc., New York, 1981, pp. 269–300.
6. J. Borges del Castillo and co-authors, *Rev. Latinoam. Quim.* **14**(3,4), 117–128 (1984).
7. E. J. Hartwell and co-workers, *J. Chem. Soc.*, 1436–1440 (1948).
8. A. Mangini and R. Passerini, *Gazz. Chim. Ital.* **87**, 243 (1957).
9. J. Behringer, *Z. Elektrochem. Angew. Phys. Chem.* **62**, 544–557 (1958).
10. K. J. Liska, and co-workers, *Tetrahedron Lett.*, 4657–4659 (1970).
11. P. Bouvier and co-workers, *Bull. Soc. Chim. Fr.*, 1187 (1977).
12. *CRC Handbook of Data on Organic Compounds*, Vol. 1, CRC Press, Inc., Boca Raton, Fla., 1985, p. 449.
13. C. V. Rechenberg, *Einfache und Fraktionierte Distillation*, S.304, Miltitz, 1923.
14. J. N. Ray, *J. Indian Chem. Soc.* **11**, 671 (1934).
15. R. Hitchens, *Ind. Eng. Chem.* **24**, 418 (1932).
16. A. Seidell, *Solubilities of Organic Compounds*, Vol. 2, 3rd ed., D. Van Nostrand Co., Inc., New York, 1941, p. 623.
17. X. A. Dominguez and co-workers, *J. Org. Chem.* **26**, 1625 (1961).
18. P. L. de Bennerville and co-workers, *J. Am. Chem. Soc.* **62**, 283 (1940).
19. Ger. Offen. 2,161,847 (June 29, 1972), R. W. Amidon and H. Greenfield (to Uniroyal, Inc.).
20. G. Chatelus, *Compt. Rend.* **224**, 201–203 (1946).
21. P. Karrer and P. Banerjea, *Helv. Chim. Acta* **32**, 1692–1693 (1949).
22. K. M. Biswas and A. H. Jackson, *J. Chem. Soc.* **12**, 1667–1674 (1970).
23. **U.S. Pat. 1,945,182 and 1,945,184** (Jan. 30, 1934), E. Clemmensen (to Monsanto Chemical Co.); F. D. Dodge, *J. Am. Chem. Soc.* **52**, 1724 (1930); **38**, 446 (1916).
24. R. C. Fuson and co-workers, *Organic Syntheses Collective*, Vol. 3, John Wiley & Sons, Inc., New York, 1955, p. 209.
25. S. M. Sethna and co-workers, *Chem. Rev.* **36**, 27 (1945).
26. **U.S. Pat. 2,478,824** (Aug. 9, 1949), H. S. Halbedel and J. C. Heath (to Harshaw Chemical Co.).
27. **U.S. Pat. 2,562,873** (July 31, 1951), W. C. Stoesser and E. H. Summerfield (to The Dow Chemical Company).
28. E. Boyland and P. Sine, *J. Chem. Soc.*, 2966–2972 (1953).
29. D. Ramesh and M. Srinivasan, *Curr. Sci.* **53**(7), 369–371 (1984).
30. F. D. Lewis and co-workers, *J. Am. Chem. Soc.* **105**(10), 3344–3345 (1983).
31. **Ger. Pat. 223,684** (Apr. 27, 1909), F. Raschig.
32. **U.S. Pat. 1,920,494** (Aug. 1, 1933), E. C. Britton and W. R. Reed (to The Dow Chemical Company).
33. **U.S. Pat. 2,062,364** (Dec. 1, 1936), G. Kretchnar and C. Collaud (to Givaudan Co.).
34. H. von Pechmann, *Berichte* **16**, 2119 (1883); **17**, 929 (1884).
35. H. Simonis, *Berichte* **48**, 1584 (1915).
36. G. C. Bailey and co-workers, *Ind. Eng. Chem.* **13**, 905 (1921).
37. A. A. Shmuk, *Vsesoyuz. Nauchn.-Issled. Inst. Tabach* **140**, 45 (1939).
38. V. G. Austerweil, *Compt. Rend.* **248**, 1810 (1959).
39. S. M. Sethna and co-workers, *Organic Reactions*, Vol. 7, John Wiley & Sons, Inc., New York, 1953, pp. 1–58.
40. R. N. Lacey, *J. Chem. Soc.*, 854 (1954).
41. T. Ziegler and co-workers, *Berichte* **120**(3), 373–378 (1987).
42. **U.S. Pat. 3,859,311** (Jan. 5, 1975), **3,936,473** (Feb. 3, 1976), T. Symon and co-workers (to Universal Oil Products Co.).
43. A. K. Das Gupta and co-workers, *J. Chem. Soc. C*, 29 (1969).
44. W. H. Perkin, *J. Chem. Soc.* **21**, 53 (1868); **31**, 388 (1877).
45. J. R. Johnson, *Organic Reactions*, Vol. 1, John Wiley & Sons, Inc., New York, 1942, p. 210.
46. H. Yanagisawa and co-workers, *J. Chem. Soc. Abstr. I* **120**, 682 (1921).
47. **U.S. Pat. 2,204,008** (June 11, 1940), E. C. Britton and co-workers (to The Dow Chemical Company).
48. K. I. Bogacheva, *Tr. Vsesoyuz. Nauchn.-Issled. Inst. Sin. Natur. Dushist. Veschestv.* (5), 30 (1961).
49. Y. Mizuno and K. Watanabe, *J. Pharm. Soc. (Japan)*, **69**, 123 (1949).
50. V. I. Isaguliants and co-workers, *J. Appl. Chem. (USSR)* **11**, 946 (1938).
51. **Ger. Pat. 3,139,994** (Apr. 21, 1983), 3,144,261 (May 19, 1983), R. Kern.
52. E. Koeppe and F. Voegtile, *Synthesis* (2), 177–179 (1987).
53. **U.S. Pat. 3,813,414** (May 28, 1974), S. Kahn (to Universal Oil Products Co.).
54. E. Knoevenagel, *Berichte* **29**, 172 (1896); **31**, 2596 (1898); **Ger. Pat. 161,171** (Dec. 1, 1903), E. Knoevenagel.
55. G. Jones, *Organic Reactions*, Vol. 15, John Wiley & Sons, Inc., New York, 1967, p. 204.
56. **Ger. Pat. 440,341** (1924), (to I. G. Farbenindustrie).
57. **Ger. Pat. 189,252** (July 27, 1905), (to Haarmann and Reimer).
58. **Jpn. Pat. 60 181,082** (Sept. 14, 1985), T. Matsuda and co-workers (to Sumitomo Chemical Co., Ltd.).
59. **Ind. Pat. 141,709 and 141,714** (Apr. 9, 1977), A. Chatterjee and co-workers.
60. **Jpn. Pat. 0 296,574** (Apr. 9, 1990), K. Okumura (to Fuso Chemical Co.).
61. **Rus. Pat. 1,030,363** (July 23, 1983), V. G. Cherkaev and co-workers (to Synth. Nat. Scent. Res.).

62. **Jpn. Pat. 3364** (Oct. 10, 1950), S. Hirokuni (to Kiyomi Chemical Industries Co.).
63. **Austrian Pat. 272,333** (July 10, 1969), H. Maier (to Oesterreichische Chemische Werke).
64. **Ger. Pat. 2,502,690** (July 31, 1975), P. Stijfs (to Stamicarbon B.V.).
65. **Ger. Pat. 1,053,119** (Mar. 19, 1959), J. A. Pettersen.
66. A. I. Goluvín and co-workers, *Vestn. Mosk. Univ. Khim.* **24**(2), 19–22 (1969).
67. G. M. Dugacheva and co-workers, *Tr. Vsesoyuz. Nauchn.-Issled. Inst. Sin. Natur. Dushist. Veshchestv.* **9**, 190–196 (1971).
68. S. V. Grijsenko and co-workers, *Farmatsiya (Moscow)* **27**(6), 45–47, (1978).
69. *Coumarin*, Rhone-Poulenc, Inc., Princeton, N. J., 1988.
70. L. W. Hazleton and co-workers, *J. Pharmac. Exp. Ther.* **118**, 348 (1956); A. J. Cohen, *Food Cosmet. Toxicol.* **17**, 277 (1979).
71. P. M. Jenner and co-workers, *Food Cosmet. Toxicol.* **2**, 327 (1964).
72. *Subchronic Oral Toxicity Test with Coumarin in Rat*, National Toxicology Program, Washington, D.C., 1981.
73. E. C. Hagan and co-workers, *Food Cosmet. Toxicol.* **5**, 141 (1967).
74. F. Griepentrog, *Toxicology* **1**, 93 (1973).
75. IARC Monograph, *Evaluation of the Carcinogenic Risks to Man*, **10**, 113 (1976).
76. A. J. Cohen, *Food Cosmet. Toxicol.* **17**, 277 (1979).
77. IARC Monograph, *Evaluation of the Carcinogenic Risks to Man, Supplement 7*, 56 (1987).
78. B. Hunter and co-workers, *Coumarin: Tumorigenic and Toxic Effects in Prolonged Dietary Administration*, Huntingdon Research Centre, personal communication, (1984).
79. I. Pekker and E. A. Shafer, *Arzneimittel-Forsch* **19**, 1744 (1969).
80. W. H. Shilling and co-workers, *Nature (London)* **221**, 664 (1969).
81. G. Feuer and co-workers, *Food Cosmet. Toxicol.* **4**, 157 (1966).
82. **U.S. Pat. 4,593,041** (June 3, 1986), J. R. Casley-Smith.
83. **Ger. Pat. 3,715,990** (Dec. 8, 1988), E. F. Elstner and co-workers (to Schaper und Bruemmer Co.).
84. **Austrian Pat. 8,503,865** (Sept. 12, 1985), J. R. Casley-Smith.
85. K. P. Link, *Fed. Proc.* **4**, 176 (1945).
86. **U.S. Pat. 2,427,578** (Sept. 16, 1947), M. A. Stahman and co-workers (to Wisconsin Alumni Research Foundation); 2,765,321 (1956) C. H. Schroeder and co-workers; 2,777,859 (1957), K. P. Link.

Paul M. Boide
Walter C. Meuly
Rhone-Poulenc, Inc.

COUMARONE–INDENE RESINS.

See [HYDROCARBON RESINS](#).

COUPLING AGENTS.

See [SILICONE COMPOUNDS](#), [SILYLATING AGENTS](#).

CRACKING.

See [CATALYSIS](#); [PETROLEUM](#).

CREAM.

See [MILK PRODUCTS](#).

CRESIDINE.

See [AMINES](#), [AROMATIC AMINES](#); [ANILINE AND ITS DERIVATIVES](#).

CRESOLS.

See [ALKYLPHENOLS](#).

CRESYLIC ACIDS, SYNTHETIC.

See [ALKYLPHENOLS](#).

CROWN ETHERS.

See [ANTIBIOTICS](#), [POLYETHERS](#); [CATALYSIS, PHASE TRANSFER](#); [CHELATING AGENTS](#).

CRYOGENICS

In present day usage, the word cryogenics refers to "all phenomena, processes, techniques or apparatus occurring or using temperatures below 120 K" (1). The temperature of a liquefied natural gas (LNG) stream at ambient pressure is generally in the neighborhood of 120 K. Refrigeration technology covers the temperature range of 120 to 273.1 K. The choice of these temperature ranges is somewhat arbitrary but has been accepted as an international standard. Around or below the temperature of 120 K, the permanent gases including methane, oxygen, argon, nitrogen, hydrogen and helium can be liquefied at ambient pressure. A few relevant thermodynamic properties for some gases at cryogenic temperatures are listed in Table 1.

Table 1. Cryogenic Properties of Gases

Property	He	Ne	Ar	H ₂	CH ₄	N ₂	O ₂	F ₂	CO
normal ^a boiling point, K	4.2	27.3	87.4	20.4	111.7	77.4	90.2	86.2	81.6
vapor density at bp, kg m ^{-3b}	16.0	9.50	5.89	1.33	1.80	4.61	4.74	4.84	4.33
liquid density at bp, kg m ^{-3b}	125.0	1200	1390	70.9	424.0	804.0	1142	1512	792
heat of vaporization at bp, J g ^{-c}	23.96	87.0	163	446.2	577.8	199	213	171	215
triple point (tp), K	2.15 ^d	19.0	83.8	14.0	88.7	63.2	54.4	53.5	61.5
vapor pressure (solid at tp), kPa ^e		43.06	68.79	7.2	10.03	12.85	0.152	0.221	3.7
heat of fusion at tp, J g ^{-c}	4.2	16.7	29.4	58.1	59.0	25.6	13.7	13.5	30.0
critical temperature, K	5.3	44.5	150.7	33.2	190.7	126.1	154.8	144.2	132.9
critical pressure, MPa ^f	0.229	2.72	4.86	1.31	4.63	3.39	5.04	5.5	3.49

^a At 101.3 kPa = 1 atm.
^b To convert kg m³ to lb ft³, multiply by 0.0624.
^c To convert J g to Btu lb, multiply by 0.429.
^d Lambda (λ) point.
^e To convert kPa to mm Hg, multiply by 7.5
^f To convert MPa to atm, multiply by 9.901.

Cryogenic technology has contributed greatly to scientific research and is widely used in many industrial applications. The ability to condense a gas such that it can be stored and shipped as a cryogenic liquid rather than a pressurized gas has found several applications. Natural gas is stored and transported as LNG to many countries in the world. Liquid hydrogen is used as fuel for space vehicles. The use of slush hydrogen, a solid–liquid mixture of hydrogen, is being considered for transonic and hypersonic aircraft. Oxygen, nitrogen, and argon are shipped as liquids by truck and rail car to the point of end use.

Another significant use of cryogenic technology has been to produce low cost, high purity gases through fractional condensation and distillation. Cryogenic air separation is used for the production of pure oxygen, nitrogen, and argon. These gases are used in primary metals manufacturing (eg, steel), chemical manufacturing, electronic industries, partial oxidation and coal gasification processes, enhanced oil recovery and many other applications. Cryogenic methods are also used for the purification of hydrogen, helium and carbon monoxide. Hydrogen and carbon monoxide gases are used in chemical manufacturing and some metal industries. Helium is used in welding, medicine, gas chromatography, and diving (2).

Cryogenic processes that provide low temperatures to refrigerate other materials or to alter their properties have been used in many applications (3). Liquid nitrogen is used for freezing food such as hamburgers and shrimp. Rubber tires and scrap metal from old cars are reclaimed using cryogenic cooling techniques to make them brittle for easier fracturing and component separation. Biological materials such as bone marrow, blood, animal semen, tissue cultures, tumor cells and skin are preserved by cryogenic freezing and storage.

Magnetic resonance imaging (MRI) uses cryogenics to cool high conductivity magnets for nonintrusive body diagnostics. Low temperature infrared detectors are utilized in astronomical telescopes. Cryogenic technology is being used to increase the speed of computers. Cryogenic refrigerators have been applied industrially for cryopumping to yield high pumping speeds and ultrahigh vacuum. With the recent advent of high temperature superconductivity, it is anticipated that applications of superconductivity at near liquid nitrogen temperature have great potential for electric power transmission, magnetic transportation systems, and magnets for energy generation in fusion processes.

Refrigeration Methods

Refrigeration for cryogenic applications is produced by absorbing or extracting heat at low temperatures and rejecting it to the atmosphere at higher temperatures. Three general methods for producing cryogenic refrigeration in large-scale commercial applications are the liquid vaporization cycle, the Joule-Thomson (J-T) expansion cycle, and the engine expansion cycle. The first two are similar in that they both utilize irreversible isenthalpic expansion of a fluid, usually through a valve. Expansion in an engine approaches reversible isentropic expansion with the performance of work.

Liquid Vaporization Cycle. In this process, a refrigerant fluid with the desired low temperature boiling point is compressed to a pressure at which it can be condensed with ambient air, cooling water or another refrigerant fluid with a higher boiling temperature. The condensed, low boiling refrigerant is isenthalpically expanded to a suitable low pressure. Evaporation of the low pressure liquid provides the desired refrigeration. The evolved vapor is then usually rewarmed prior to recompression. Heat rejection can be cascaded from very low temperatures to ambient level by use of several refrigerants with different boiling temperature characteristics.

Joule-Thomson (J-T) Expansion Cycle. In this process, a refrigerant fluid is compressed and precooled below its inversion temperature, ie, the temperature below which a pressure reduction results in a temperature decrease. The cold refrigerant fluid is isenthalpically (J-T) expanded to a lower pressure to produce the required low temperature. The low temperature fluid is partially rewarmed to provide cryogenic refrigeration, and is usually then further warmed against the high pressure refrigerant fluid to provide the precooling.

Engine Expansion Cycle. In this process, a refrigerant fluid is compressed, optionally precooled, and expanded to a lower pressure through an expander while producing work to reduce both the enthalpy and temperature of the fluid. The expanded refrigerant fluid is warmed to provide cryogenic refrigeration.

Other Refrigeration Methods. Cryocoolers provide low temperature refrigeration on a smaller scale by a variety of thermodynamic cycles. The Stirling cycle follows a path of isothermal compression, heat transfer to a regenerator matrix at constant volume, isothermal expansion with heat transfer from the external load at the refrigerator temperature, and finally heat transfer to the fluid from the regenerator at constant volume.

The Gifford-McMahon cryocooler consists of displacer, regenerator, compressor and intake exhaust valves that can be staged to reach cryogenic temperatures.

Magnetic refrigeration uses the magnetocaloric effect to produce cooling. When ferromagnetic or paramagnetic materials are put into a magnetic field, their temperature rises: when removed from the field, their temperature falls. The amount of temperature change depends on the material, its temperature, and the strength of the magnetic field.

Applications

Air Separation. A considerable number of cryogenic concepts were developed near the end of the 19th and beginning of the 20th century to liquefy and separate air into its main constituents: 78% nitrogen, 21% oxygen and 1% argon by volume. By 1905, Carl von Linde's double distillation column process to produce practically pure oxygen and nitrogen was already known. Argon was industrially produced by 1912 (4). Trace quantities of other rare gases such as neon, krypton, and xenon are present in air and can be recovered by proper modifications to a cryogenic double-distillation column air-separation plant (2).

Figure 1 shows a cryogenic air-separation process using a double-distillation column of the Linde-type with a side rectifying column to recover crude argon (5). Feed air is compressed in a multistage compressor to about 650 kPa (94 psi) with interstage cooling. The compressed air is cooled in an aftercooler by cooling water and the condensed water is removed in an aftercooler separator. To avoid freezing of water and carbon dioxide in the cryogenic part of the plant, the feed air is passed through an adsorbent bed of molecular sieves. Along with water and carbon dioxide, several trace impurities in the air such as acetylene, ethylene, butane, and other heavier hydrocarbons are also adsorbed. This alleviates some of the safety problems arising from the formation of hydrocarbon–oxygen mixtures in older plants that did not use molecular-sieve adsorbents.

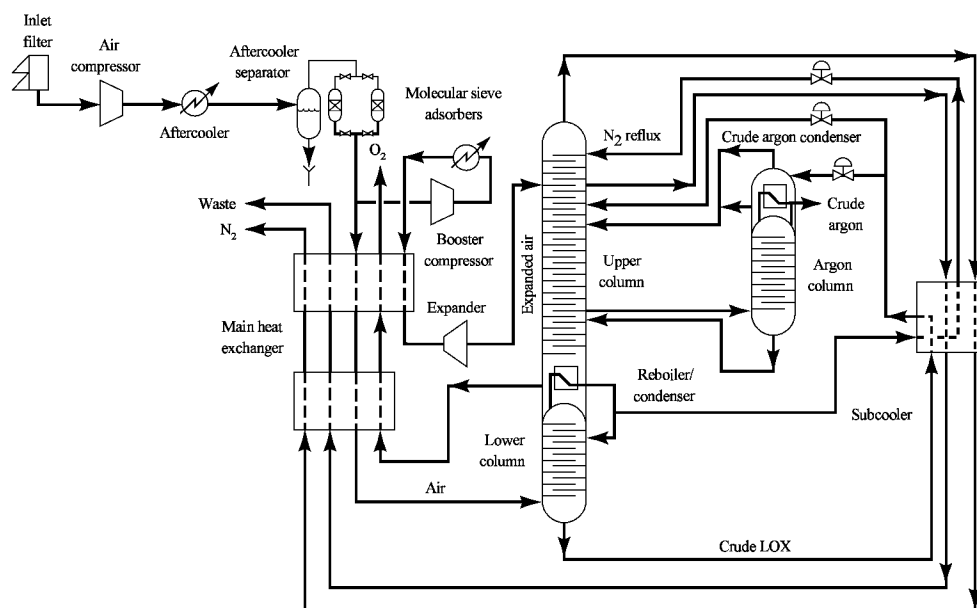


Fig. 1. Cryogenic air separation process. LOX = liquid oxygen.

The purified air stream is split into two streams. One stream of about 7–12% of the total air flow is boosted in pressure by about 60–100 kPa (9–15 psi) in a booster compressor. The boosted air is cooled by cooling water and then in the main heat exchanger against the returning cold product nitrogen and oxygen streams. The cooled air stream is work expanded to near atmospheric pressure and fed to the distillation column. This expander provides the needed refrigeration for the plant. Usually the work extracted from the expander drives the booster compressor such that an additional source of energy is not required.

The principal air stream is cooled in the main heat exchanger to near its dew point. Nitrogen and oxygen are separated by distillation in a two-stage distillation process. The first stage, the lower column, which operates at about 600 kPa (87 psi) separates the feed air into nitrogen and an oxygen-enriched liquid (crude LOX) stream, which is eventually fed to the upper column. The upper column, which operates at close to ambient pressure, produces pure oxygen and nitrogen streams. Nitrogen reflux for both columns is generated at the top of the lower column. Nitrogen vapor at the top of the lower column is condensed against the liquid oxygen at the bottom of the upper column by heat exchange between the two streams in a reboiler-condenser. The pressure difference between the two columns provides the temperature difference between the condensing and boiling fluids.

Since argon boils between oxygen and nitrogen, a peak in the argon concentration of about 7–15% occurs in the lower section of the upper column. A vapor stream is drawn from the upper column near the location of the peak argon concentration and is distilled in a third (argon) column. Typically, a vapor stream is drawn from the upper column at a point where the concentration of nitrogen is low (25–100 vppm) and the concentration of argon is 7–12%. Reflux for the argon column is provided by heat exchange between the vapor at the top of this column and a portion of the crude LOX stream in a crude argon condenser. The vaporized crude LOX is fed to the upper column and the condensed argon-rich stream is returned as reflux to the argon column. The liquid from the bottom of the argon column is returned to an appropriate location in the upper column. A crude argon stream containing 2–5% O₂ and 0.05–1% N₂ is withdrawn from the top of the argon column. The oxygen produced from the bottom of the upper column can easily have an oxygen purity of 99.5% or greater with the remainder being mainly argon. The nitrogen product from the top of the upper column can be produced with an oxygen concentration below 5 ppm by volume (vppm).

The heat exchangers used in cryogenic air-separation plants are generally brazed aluminum plate-fin heat exchangers. Since the early 1930s, sieve trays have been used predominantly for cryogenic distillation. Recently, the use of structured packing with much lower pressure drop per theoretical stage of separation is becoming increasingly common in the upper and argon columns. Its use in the upper column decreases the power consumption by the air compressor as the pressure of the feed air to the lower distillation column can be decreased by about 60 kPa (9 psi). The use of structured packing in the argon column provides more theoretical stages of separation for the same permissible pressure drop between the bottom and top of the argon column. This allows argon production from the argon distillation column with oxygen impurity as low as 1 vppm. This either eliminates or reduces the size of the expensive downstream processes that have been used in the past to produce pure argon product (6).

In the field of semiconductor device manufacturing, ongoing trends toward device miniaturization and the need for high production yield are leading to gas specifications with ultra-high purities. This requires that concentration of all impurities in each of the nitrogen, oxygen, and argon products be less than 10 vppb. It is widely felt that this impetus to decrease the impurity level will continue as the devices continue to advance and gases with parts-per-trillion impurity concentrations will shortly be needed. To meet this recent challenge, cryogenic processes modifying the scheme shown in Figure 1 by employing additional columns have been developed. For example, use of an additional stripping column in conjunction with the upper column to coproduce a large fraction of oxygen product as ultrahigh purity oxygen has been described (7).

Liquid Nitrogen, Liquid Oxygen, and Liquid Argon. Large quantities of liquid nitrogen are typically produced by liquefying gaseous nitrogen from an air separation plant. For this purpose, a gaseous nitrogen stream is compressed to a fairly high pressure and work expanded at successive lower temperatures to create colder temperatures for liquefying nitrogen. Figure 2 shows one such process (1). Low pressure gaseous nitrogen from the air-separation unit is combined with a recycle nitrogen stream and compressed to about 600 kPa (87 psi) in a feed compressor. The compressed nitrogen is combined with another recycle nitrogen stream and further compressed to about 3 MPa (435 psi) in a recycle compressor. Both the feed and recycle compressors are high efficiency, multistage centrifugal machines. The nitrogen stream is then further compressed in two booster compressors to about 4.5 MPa (653 psi), which is above the critical pressure of nitrogen. Each of these booster compressors is linked to and driven by an expander providing refrigeration to the process. The final compressed nitrogen stream is then sent to a heat exchanger. Large portions of this stream are expanded in two expanders and recycled, while a smaller fraction is cooled to below the critical temperature of nitrogen by heat exchange with the expanded streams. The inlet temperatures to the expanders are chosen such that the temperature differences between the cooling and warming streams in the heat exchanger are minimized to reduce the liquefaction power requirement. The cooled high pressure nitrogen stream is then isenthalpically let down in pressure across valves in two stages to provide the desired liquid nitrogen product stream near ambient pressure.

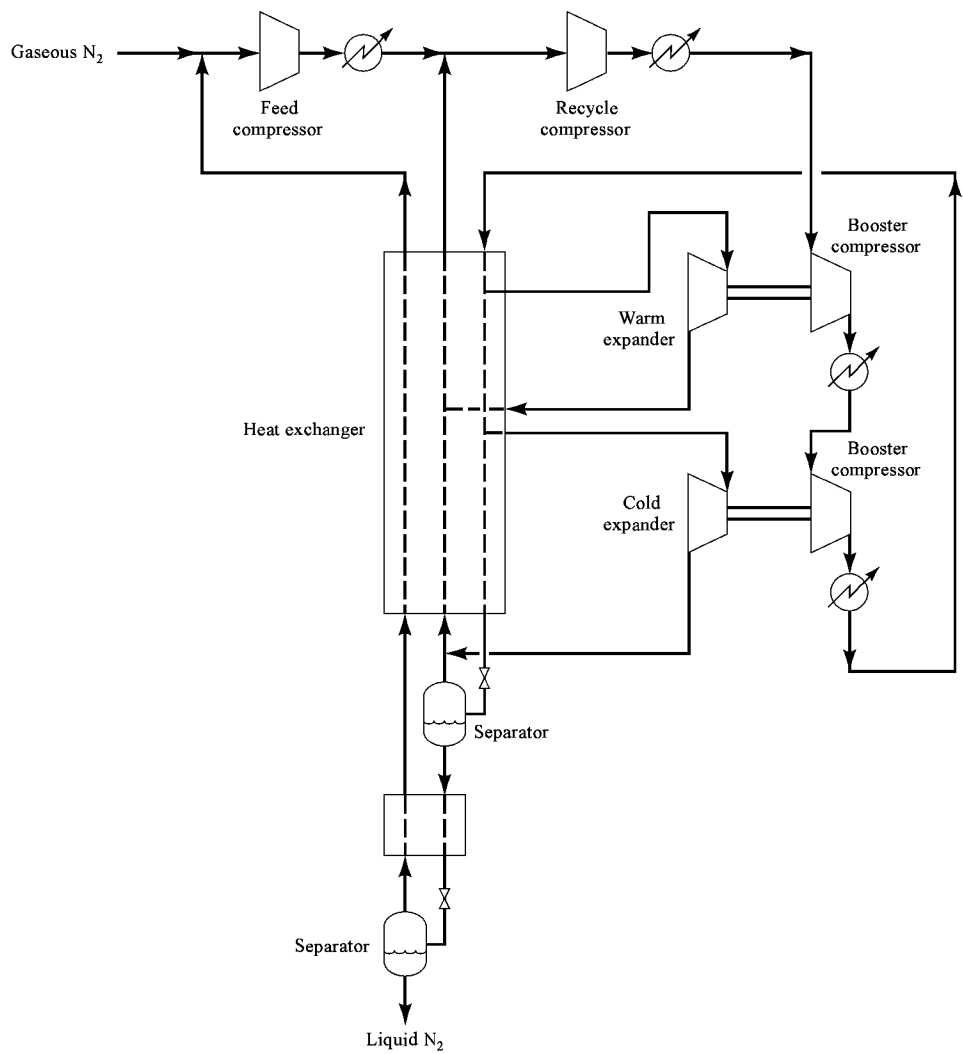


Fig. 2. Nitrogen liquefaction process.

Liquid argon can be produced directly from the crude argon condenser shown in Figure 1 by increasing the flow of air through the expander, but expander refrigeration is not economic for production of large quantities of liquid oxygen. Typically, for this purpose, refrigeration is provided by adding liquid nitrogen from a nitrogen liquefier to the top of the upper column while liquid oxygen is withdrawn from the bottom of this column.

Liquefied Natural Gas.

Liquefied Natural Gas (LNG) plants can be categorized as peakshaving or baseload. Peakshaving LNG plants are built at the consumer end of natural gas pipelines to accumulate LNG in storage tanks for later vaporization and sendout into the local grid during periods of peak demand. Baseload LNG plants provide a steady base supply of natural gas to utility companies, generally by transportation of LNG by ship from one country to another.

Peakshaving LNG plants shave the high peaks of natural gas production and transmission by putting vaporized LNG back into the local grid during high demand periods. Peakshaving plants liquefy natural gas during the 125 to 200 day season of the year when excess supply pipeline capacity exists. Then, during cold weather periods of high demand, the LNG is pumped to distribution pressure, vaporized, and put into the local grid. This allows a higher yearly utilization of gas without expanding the production and transmission facilities. Peakshaving LNG plants range in size up to 540,000 STP m³ per day, have LNG storage sufficient to hold one season's production, and sendout capacity (pumps and vaporizers) to return the season's production to the grid during 5 to 15 days of cold weather.

Pretreatment of the natural gas before liquefaction is necessary to remove acid gases (CO₂ and H₂S) and water, which would freeze and plug the liquefaction equipment. The gas to peakshaving plants has been treated to pipeline specifications for acid gases and water before transmission. At the peakshaving plant, residual acid gases are removed by absorption with monoethanolamine [141-43-5] (MEA), methyldiethanolamine [105-59-9] (MDEA), or other amine solutions. If there is sufficient regeneration gas available and the acid gas content is low enough, CO₂ and H₂S can be co-adsorbed on beds of molecular sieve along with water. For baseload LNG plants, the acid gas content may be very high. MEA, MDEA, Sulfinol, Benfield or other solutions are used, depending on the CO₂ H₂S partial pressures, the required purity, and economics.

The most commonly used refrigeration cycles for LNG peakshaving plants are the cascade J-T expansion cycle, the expander cycle, and the multicomponent refrigerant cycle. The cascade J-T expansion cycle typically involves three separate working fluids, propane, ethylene, and methane, in individual refrigeration loops. Propane is compressed to a pressure at which it can be totally condensed against cooling water or air. The propane may be vaporized at several pressures while cooling the natural gas feed, ethylene, and methane streams. Ethylene is compressed to a pressure sufficient to be condensed by propane vaporizing in its lowest stage at a pressure near one atmosphere. Likewise methane is condensed against the lowest boiling stage of ethylene, which may also have several intermediate pressure (temperature) levels. Staging each of the propane, ethylene, and methane circuits to have intermediate vaporizing pressures reduces the refrigeration power requirement for LNG production by reducing the compression ratios of the refrigerant streams.

Expander cycles provide refrigeration by having a fluid perform work in an expansion engine. The fluid can be nitrogen or natural gas. One variation of the expander cycle uses a relatively large quantity of available flowby gas that is normally throttled from transmission pressure to distribution pressure. Rather than reducing pressure across valves, the flowby natural gas is precooled and work expanded to provide refrigeration to liquefy a small portion of the natural gas. The flowby gas is then rewarmed and enters the distribution grid. Using this flowby gas reduces both capital and operating costs of the peakshaving plant.

The multicomponent (mixed) refrigerant cycle utilizes J-T refrigeration from a mixture including nitrogen and a blend of hydrocarbons from methane to butane or pentane. This cycle has a simplified arrangement compared to the other cycles. The mixture composition is optimized to obtain close temperature approaches to the condensing feed gas, and reduce the power for LNG production. The mixture is compressed to about 3 MPa (435 psi), partially condensed against cooling water, precooled, condensed, and subcooled in a heat exchanger, then reduced in pressure across a J-T valve to about

0.3 MPa (44 psi) for vaporizing and rewarming to cool the natural gas feed and high-pressure refrigerant.

Baseload LNG plants have predominantly used variations of the mixed refrigerant cycle, although some have utilized the cascade cycle. The mixed refrigerant cycle can be precooled by a high stage cycle that may be a propane refrigerant, a second separate mixed refrigerant cycle, or even an ammonia absorption system (8–11). The propane precooled mixed refrigerant cycle shown in Figure 3 has low energy requirements, characteristic of the cascade cycle, but a simpler equipment arrangement. Baseload plants generally recover propane and heavier components from the natural gas for internal use in charging the refrigeration systems, or for external use as LPG, butane, or natural gasoline. Baseload LNG plants range in size from about 3 to 10 million STP m³ per day LNG production. Machinery for the baseload plants is centrifugal or axial compressors driven by steam or gas turbines. Baseload LNG plants serve the trade to Japan, Taiwan, and Korea from Alaska, Malaysia, Borneo, Indonesia, Australia, and the Persian Gulf, as well as trade from North Africa to Europe and the United States.

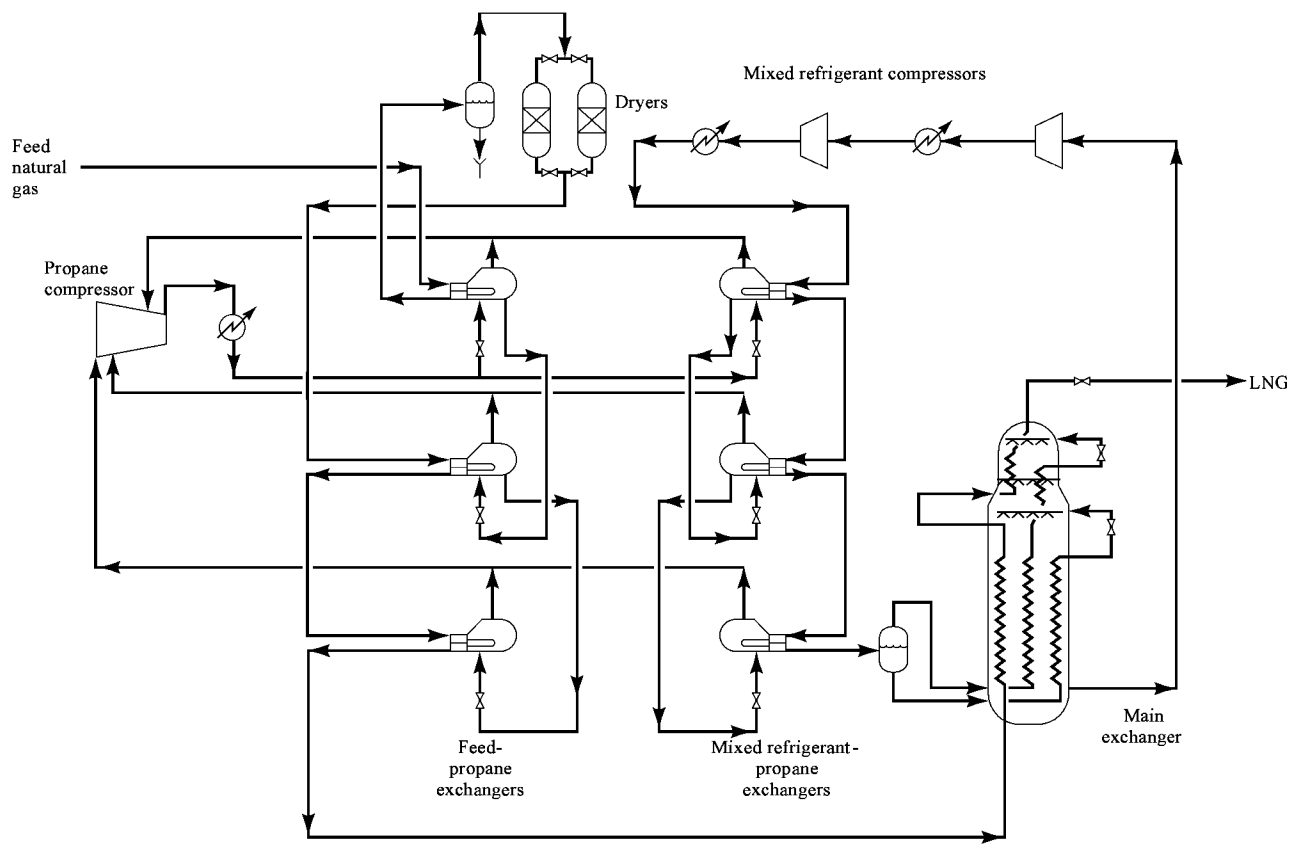


Fig. 3. Propane precooled mixed refrigerant process for natural gas liquefaction.

Hydrogen Purification.

Cryogenic separation is used extensively for recovery and purification of hydrogen from refinery and petrochemical plant off-gases. The most common applications are in recovery of hydrogen from catalytic reformer gas, ethylene plant off-gas, hydrodealkylation (HDA) recycle gas, hydrodesulfurization (HDS) off-gas and methanol–ammonia synthesis purge gases (12). Hydrogen can also be coproduced when carbon monoxide is recovered by cryogenic separation of a steam–methane reformer off-gas.

The simplest form of hydrogen purification uses partial condensation to condense light hydrocarbons and trace impurities such as argon, carbon monoxide, and nitrogen from the relatively noncondensable hydrogen. Because of the high relative volatility of hydrogen to light hydrocarbons, in the range of 100 to 300 for hydrogen–methane at separation temperatures of 135–100 K, cryogenic separation can easily attain a hydrogen purity of 90–96% without distillation. Product hydrogen is recovered at close to feed gas pressure, with typical hydrogen recoveries of 90–98%.

At optimum conditions, ie, feed gas rate of at least 5000 STP m³ per hour, feed pressure of at least 2.8MPa (406 psi) and feed hydrogen content less than 80%, all of the refrigeration requirements for the cryogenic hydrogen purification process can usually be obtained by Joule-Thomson expansion of the condensed hydrocarbons. At less favorable conditions, auxiliary refrigeration is required. This refrigeration can typically be supplied by inexpensive package Freon or propane units providing refrigeration at temperatures of 250 to 210 K.

Compact brazed aluminum plate-fin heat exchangers can be used in most cryogenic hydrogen purification applications. The use of these relatively low cost heat exchangers, combined with low separation energy requirements, results in a highly economical process for hydrogen purification.

One of the most common applications of cryogenic hydrogen purification is to recover hydrogen and reject light hydrocarbons from the recycle loop of an HDA unit that converts toluene or other aromatics to benzene (Fig. 4). After pretreatment of the feed gas to remove water and aromatic impurities that would freeze at low temperatures, the feed gas is cooled to about 120 K to condense most of the hydrocarbons and provide a product hydrogen purity of 90–92%. The condensed hydrocarbons are flashed to a low pressure, typically 200–500 kPa (29–73 psi) revaporized and warmed for refrigeration recovery and then sent to the plant fuel system. The product hydrogen is rewarmed at pressure and recycled to the HDA unit.

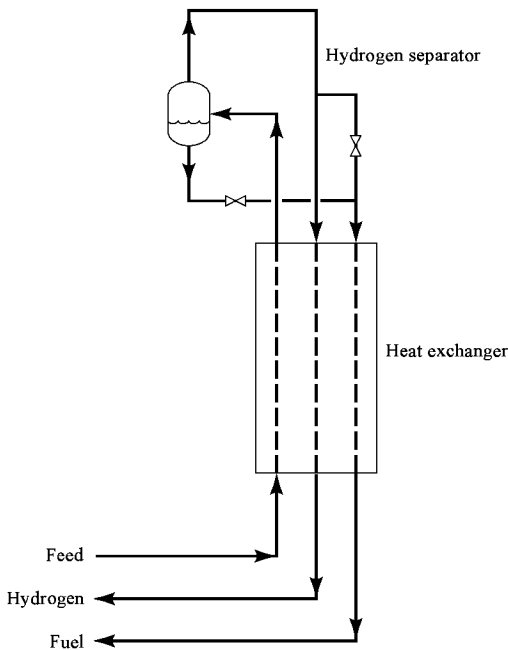


Fig. 4. Hydrogen purification process.

Hydrogen Purification with Light Olefins and Liquefied Petroleum Gas (LPG) Recovery. The relatively simple cryogenic purification process for hydrogen recovery can easily be adapted (13) to recover a crude light olefin or propane and heavier hydrocarbon (LPG) stream (Fig. 5). The pretreated feed gas is cooled to an intermediate temperature, in the range of 240 to 200 K for propylene LPG recovery or 180 to 150 K for ethylene recovery. The uncondensed vapor is then further cooled to condense the remaining methane and residual heavy hydrocarbons. The two condensed hydrocarbon streams are flashed separately to a lower pressure, revaporized and warmed with the product hydrogen stream for refrigeration recovery. A flash separator can be added to reduce the amount of light impurities in the crude olefin LPG stream. A demethanizer or de-ethanizer column can also be incorporated to provide a high purity olefin or LPG product, but will require auxiliary refrigeration.

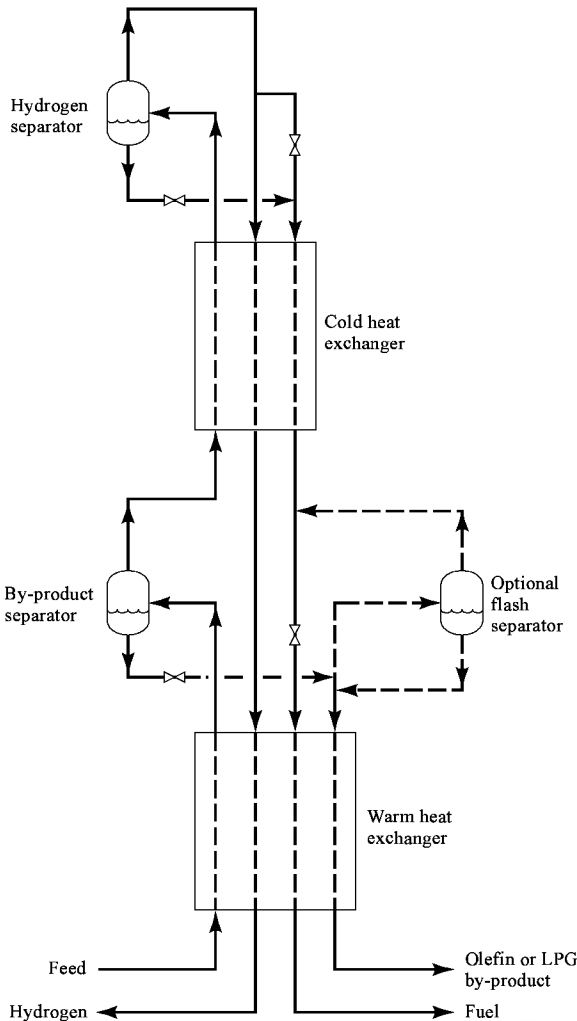


Fig. 5. Hydrogen purification process with by-product recovery.

Hydrogen Liquefaction. Hydrogen can be produced from caustic–chlorine electrolytic cells, by decomposition of ammonia or methanol, or by steam–methane reforming. Hydrogen recovered by these methods must be further purified prior to liquefaction. This is generally achieved by utilizing pressure swing adsorption methods whereby impurities are adsorbed on a solid adsorbent. Except for neon, there are no gases boiling between nitrogen and hydrogen, so cascading of refrigeration by condensing one fluid against another boiling at above ambient pressure is not feasible. Consequently, refrigeration below 77 K is provided by expanders, or by vacuum nitrogen (triple point 63.2 K) in combination with expanders. A cycle utilizing only hydrogen expanders for refrigeration (no liquid nitrogen precooling) is shown in Figure 6 (14–15).

As temperatures decrease, closer temperature approaches are needed in the heat exchangers to achieve low energy requirements. Consequently, temperature pinches in liquid hydrogen plants range from 1 degree K at 20 K to 6 degrees K at 300 K.

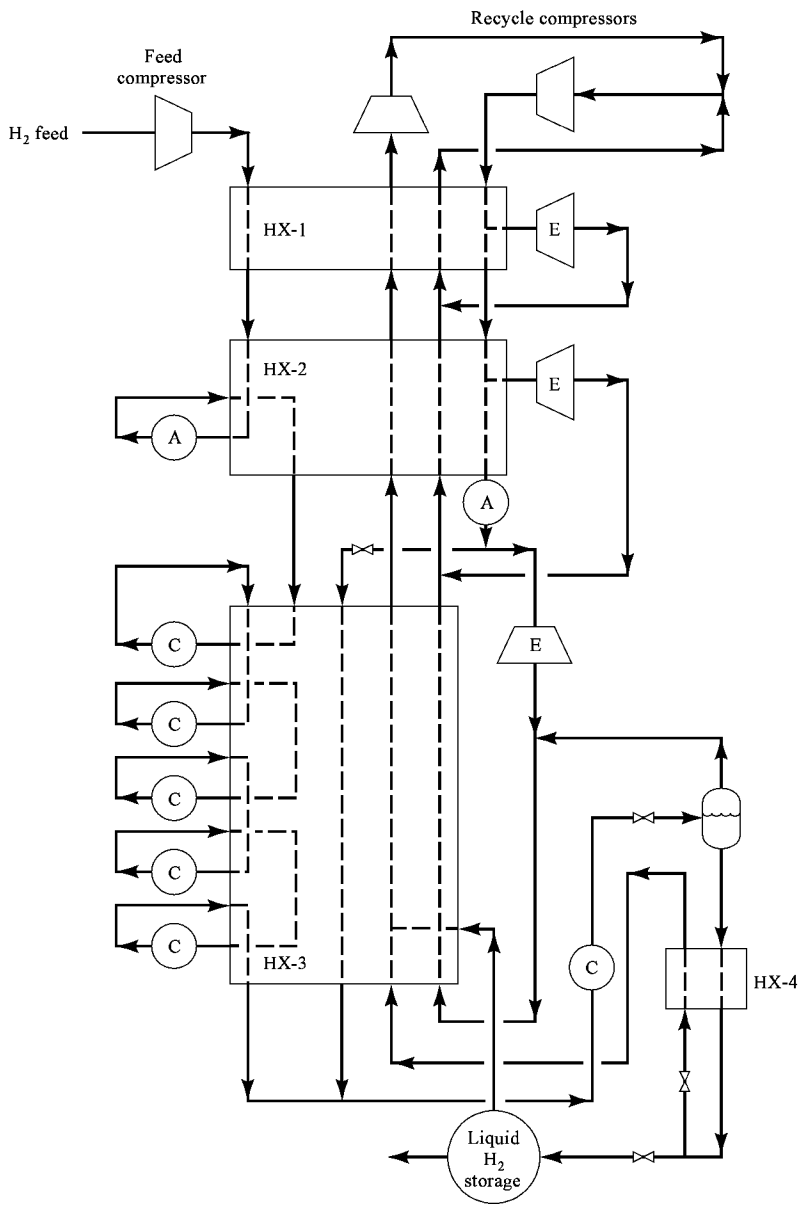


Fig. 6. Hydrogen liquefaction process. A represents adsorber; C, catalyst bed; HX, heat exchanger; E, expander.

Hydrogen has two co-existing isomers (ortho and para) that are present at equilibrium in a temperature-dependent ratio (16). The equilibrium para content of hydrogen is 25% at high temperature, 50.5% at 77 K, and 99.8% at 20.2 K. The ortho-to-para reaction is reversible, exothermic in the ortho-to-para (o → p) direction, and endothermic in the para-to-ortho (p → o) direction. Normal (25% para) hydrogen can be liquefied, but the conversion from ortho to para occurs in liquid hydrogen even without a catalyst. The heat of the o → p reaction causes a large amount of the liquid to boil away, due to the high heat of conversion, 702.6 J/g converted (168 cal/g) relative to the latent heat, 445.5 J/g (106.5 cal/g) (17). Having the heat of conversion enter the process at the lowest temperature is thermodynamically undesirable and results in high liquefaction energy. To avoid this and thereby reduce energy requirements, catalyst beds are used at various levels in the process so that the o → p reaction (with its heat of conversion) occurs at as high a temperature as possible, given the constraint of equilibrium. Hydrogen may contact catalyst continuously as it is cooled or in individual stages (14–15,18). Liquid hydrogen is usually produced as 95% para as an economic balance between the costs of liquid boil-off (power) and large catalytic converters.

The U.S. military specification, MIL-P-27201B, requires 95% para content, 99.995% minimum hydrogen by difference, 50 vppm maximum total impurities, 9 vppm maximum combined nitrogen, water, and volatile hydrocarbons, 1 vppm maximum combined oxygen and argon, 39 vppm maximum helium, 1 vppm maximum carbon monoxide and dioxide, and a 10–40 micrometers nominal absolute particulate filtration level. Liquid hydrogen is stored in double-walled vessels with evacuated pedite or multilayer insulation and transported in similarly insulated 50,000-L trailers or 900,000-L barges.

Slush hydrogen is a mixture of liquid and solid hydrogen at the 13.8 K triple point (19). A 50% solid slush mixture can be pumped through piping, valves, and flow meters without compacting. Fifty percent slush is being considered as a fuel for space vehicles because of its greater density relative to normal boiling point liquid, 81.41 kg/m³ vs 70.71 kg/m³, and its greater enthalpy change to normal boiling point vapor compared to that for liquid, 527.26 J/g (126 cal/g) vs 445.5 J/g. It is made by the vacuum freeze-thaw method which oscillates the pressure about the triple-point pressure (7.04 kPa ≈ 1 psi) while stirring to disperse the solids, or by the auger method which uses a scraped-wall heat exchanger with helium refrigeration to freeze the hydrogen that the auger then scrapes from the wall. The auger method can operate at a positive pressure. Safety measures must be taken to avoid air being drawn into the vacuum hydrogen to avoid a hazardous accumulation of oxygen.

Light Olefins and LPG Recovery. Even though the normal boiling point temperature of ethylene (169.4 K) is much above 120 K, its recovery often requires much lower processing temperatures, particularly when high recoveries are needed.

As described above, light olefins and LPG can be recovered by making proper modifications to a hydrogen recovery process. Dephlegmators, or fractionating condensers, can also be used for light olefin and LPG recovery (20–21). Dephlegmators are usually specially designed brazed aluminum plate-fin heat exchangers in which liquid condensed from an upward cooling vapor stream flows countercurrently within the heat exchanger to act as a reflux stream. The resulting mass transfer between the liquid and vapor, typically equivalent to 5–15 theoretical stages, produces a much higher purity liquid product than the conventional partial condensation process. Since less light components are liquefied, less refrigeration is required, and higher recovery of the olefin/LPG products is economically justified.

A dephlegmator process can be used to recover ethylene–ethane and heavier hydrocarbons from fluid catalytic cracking (FCC) unit off-gas (Fig. 7). Pretreated feed gas is cooled to about 230 K and then further cooled and rectified in a dephlegmator to recover 90 to 98% of the ethylene, 99% of the

ethane and 100% of the heavier hydrocarbons. Refrigeration for the cryogenic separation process is provided by revaporizing and warming the concentrated liquid C2+ hydrocarbon stream at low pressure and by work expanding the rejected light gases to fuel system pressure. The work extracted from this fuel expander can be used to drive a fuel compressor to recompress the warm fuel stream to a higher pressure.

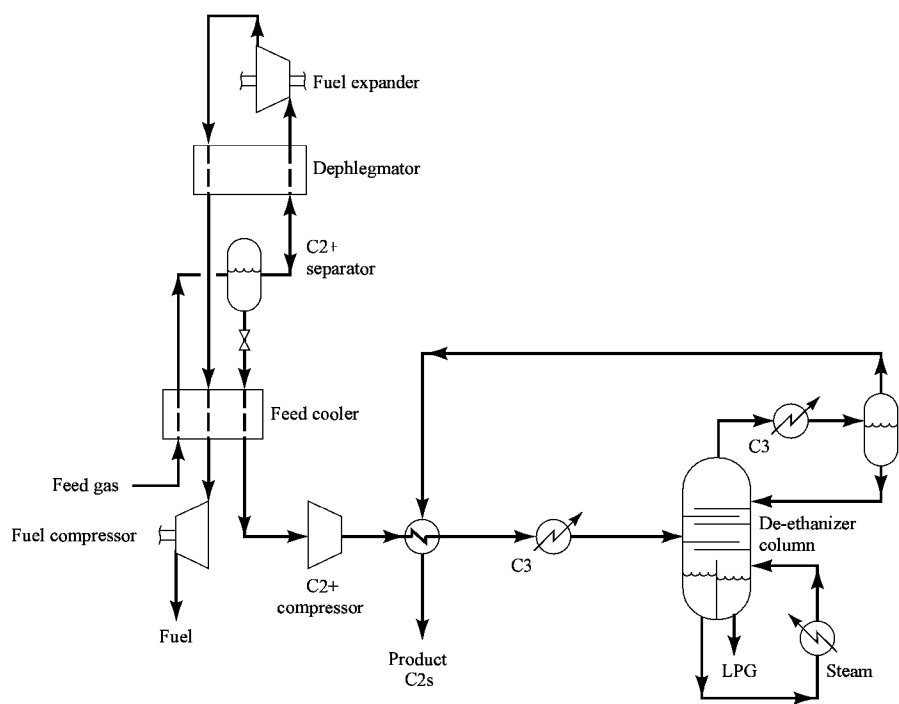


Fig. 7. FCC off-gas ethane-ethylene recovery process. C3 = propane refrigerant.

The dephlegmator process recovers a substantially higher purity C2+ hydrocarbon product with 50–75% lower methane content than the conventional partial condensation process. The C2+ product from the cryogenic separation process can be compressed and further separated in a de-ethanizer column to provide a high purity C3+ (LPG) product and a mixed ethylene-ethane product with 10–15% methane. Additional refrigeration for the deethanization process can be provided by a package Freon, propane or propylene refrigeration system.

Nitrogen Rejection and Helium Recovery.

Cryogenic distillation has been used extensively in the processing of natural gas for nitrogen removal and for helium recovery (22–23). Two basic processes are now used for nitrogen rejection from natural gas—the single-column heat-pumped process and the double-column process. Earlier processes utilized multistage flash columns for helium recovery from natural gas (24).

In the single-column heat-pumped process (Fig. 8), feed gas is cooled and fed to a high pressure column which operates at 2.1–2.8 MPa (304–406 psi). Nitrogen vapor is withdrawn from the overhead of the column and natural gas liquids from the bottom. A closed-loop methane heat pump supplies reboiler heat and condenser cooling to effect the separation. The natural gas liquids stream is flashed to a low pressure, revaporized, and warmed with the rejected nitrogen for refrigeration recovery. The nitrogen rejected at high pressure is suitable for reinjection into oil reservoirs to maintain pressure and enhance oil recovery.

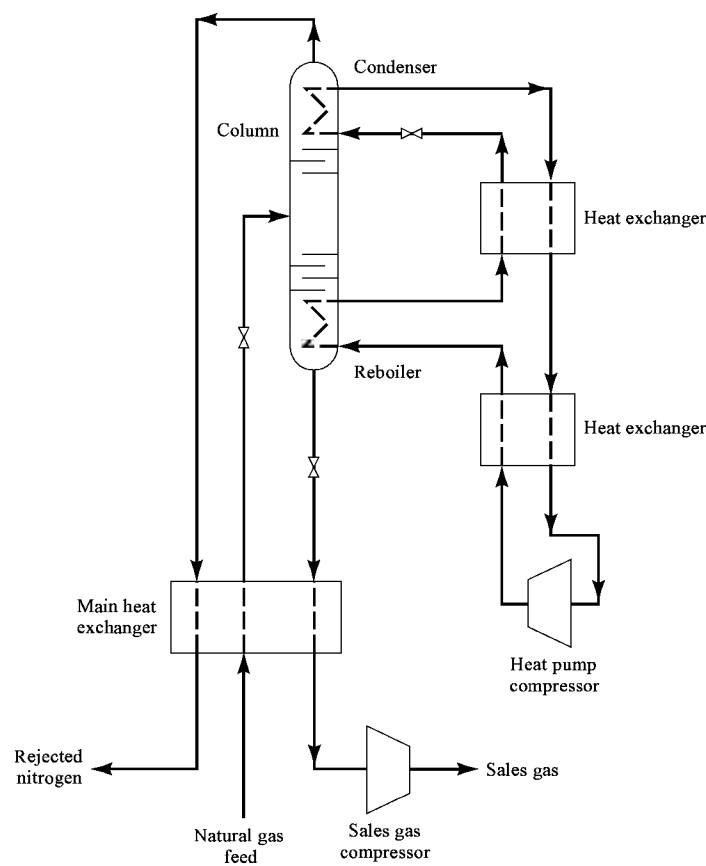


Fig. 8. Single-column heat pumped process for nitrogen rejection.

The double-column process (Fig. 9) is similar to the process described in Figure 1 for air separation. Feed gas is cooled and fed to a high pressure column operating at 1–2.5 MPa (145–363 psi). Nitrogen from the top of the high pressure column and a crude natural gas liquid stream from the bottom of the high pressure column are further processed in a low pressure column at about 150 kPa (22 psia) to complete the separation. Reboiler heat for the low pressure column is provided by condensing nitrogen at the top of the high pressure column, which is then used to reflux both columns. The natural gas liquids from the bottom of the low pressure column are pumped to a higher pressure, revaporized, and warmed with the low pressure nitrogen from the top of the low pressure column for refrigeration recovery.

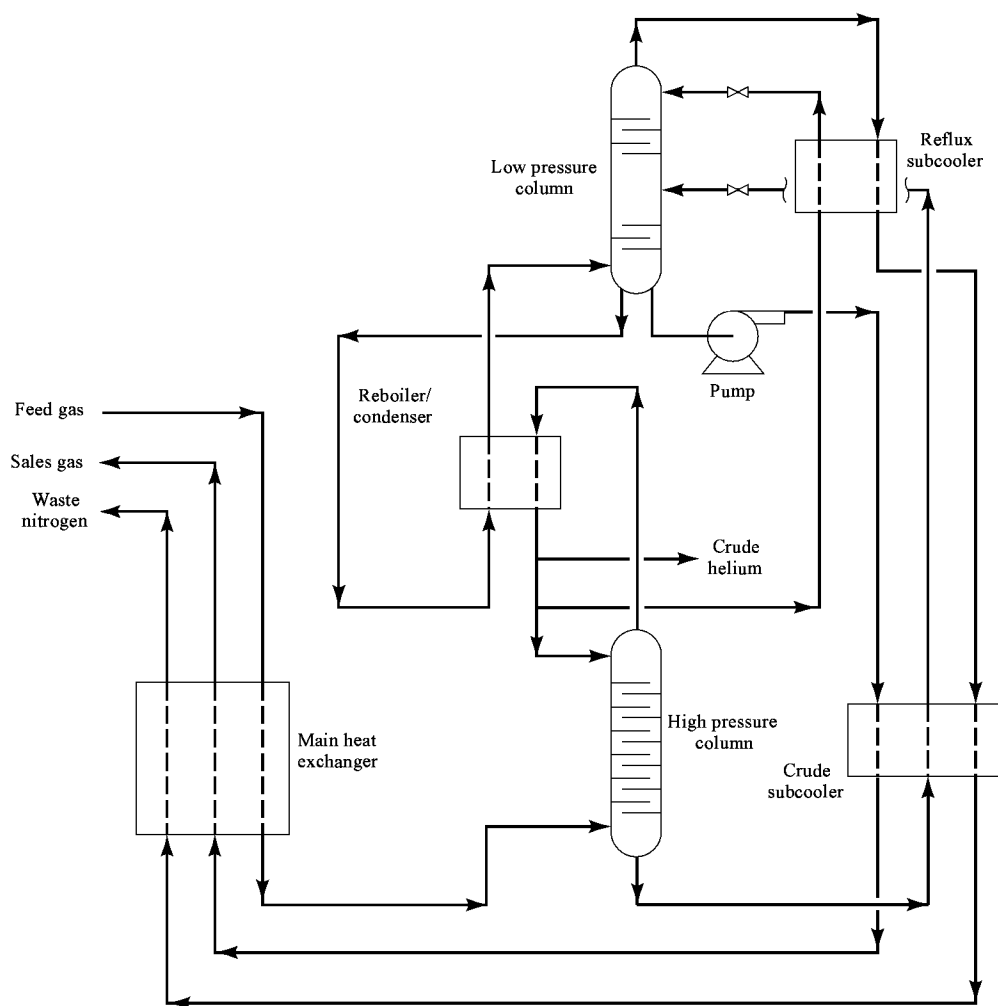


Fig. 9. Double-column process for nitrogen rejection.

Crude helium (containing 50–70% helium, associated hydrogen and neon, 1–3% methane, and the balance nitrogen) can easily be obtained by minor enhancements to the nitrogen rejection unit, particularly with natural gases containing 0.5% or more helium. For example, by operating the double-column condenser in a partial condensation mode, a stream of uncondensed vapor at about 50% helium concentration can be obtained. This crude helium stream can be fed directly to helium purification and liquefaction units.

Natural gas liquids (NGL) recovery can also be integrated into a nitrogen rejection unit to achieve economies in equipment and operating cost (25). Integration of the heat exchange equipment and optimization of the refrigeration in the nitrogen rejection and NGL sections of the plant can provide significant overall savings. Typical ethane recoveries of 70–85% can be obtained.

Helium Purification and Liquefaction. Helium, which is the lowest-boiling gas, has only 1 degree K difference between its normal boiling point (4.2 K) and its critical temperature (5.2 K), and has no classical triple point (26,27). It exhibits a phase transition at its lambda line (running from 2.18 K at 5.03 kPa (0.73 psia) to 1.76 K at 3.01 MPa (437 psia)) below which it exhibits superfluid properties (27).

Helium is commercially recovered from natural gas. If helium is present in a natural gas it usually occurs in concentrations between 0.2% to 2%, although it has been found in concentrations up to 8%. In the United States, significant quantities of helium exist in gas fields of Wyoming and the panhandle regions of Texas and Oklahoma and in southwestern Kansas. Helium is also found in natural gas in Algeria, Canada, Poland, China, and the North Sea. As part of the United States' helium conservation program of the 1960s, several plants were built to recover helium (typically processing 12 million STP m³ per day of gas), as well as a 400 km pipeline to collect their crude helium. Crude helium from this pipeline is stored in the Cliffside Reservoir near Amarillo, Texas for later withdrawal and processing. Helium from the Wyoming gas is liquefied without intermediate storage as crude helium.

Helium is normally concentrated in stages, first from the initial field concentration to crude, then to pure helium for liquefaction or sale as pure gas. As shown in the previous section, crude helium is easily recovered from a plant rejecting nitrogen from natural gas. Final upgrading of helium from crude to pure is typically done by Pressure Swing Adsorption (PSA) in combination with cryogenic partial condensation (28) (see ADSORPTION, GAS SEPARATION). Helium in natural gas streams is accompanied by neon and hydrogen that cannot be separated from helium in the PSA unit; however, both must be removed to prevent their freezing in the helium liquefaction process.

A process for final upgrading (purifying) of the helium is shown in Figure 10 (28). Crude helium, which may be at low temperature from previous processing, is combined with PSA purge gas recycle, cooled to 80 K and partially condensed to provide a vapor stream of about 90% helium concentration. Liquid from the partial condensation is warmed and separated to provide a waste stream and a nitrogen stream. The waste stream is used to regenerate the driers, and a part of the nitrogen stream may be liquefied. The upgraded 90% helium stream, with air added to provide oxygen for catalytic combustion of the hydrogen, is preheated and passed over a hydrogen removal catalyst, then cooled to condense most of the water (Fig. 11). The upgraded helium stream then flows through the PSA unit and to liquefaction. The PSA unit removes most of the remaining impurities (including water and nitrogen but excluding neon and unreacted hydrogen) to less than 10 vppm. The trace impurities are removed in an 80 K adsorber in the liquefier, and neon and any unreacted hydrogen are adsorbed in a 20 K adsorber. A portion of the pure product from the PSA is used to regenerate the PSA beds, then compressed, dried, and recycled to the inlet of the upgrader where it is combined with crude feed for high overall helium recovery.

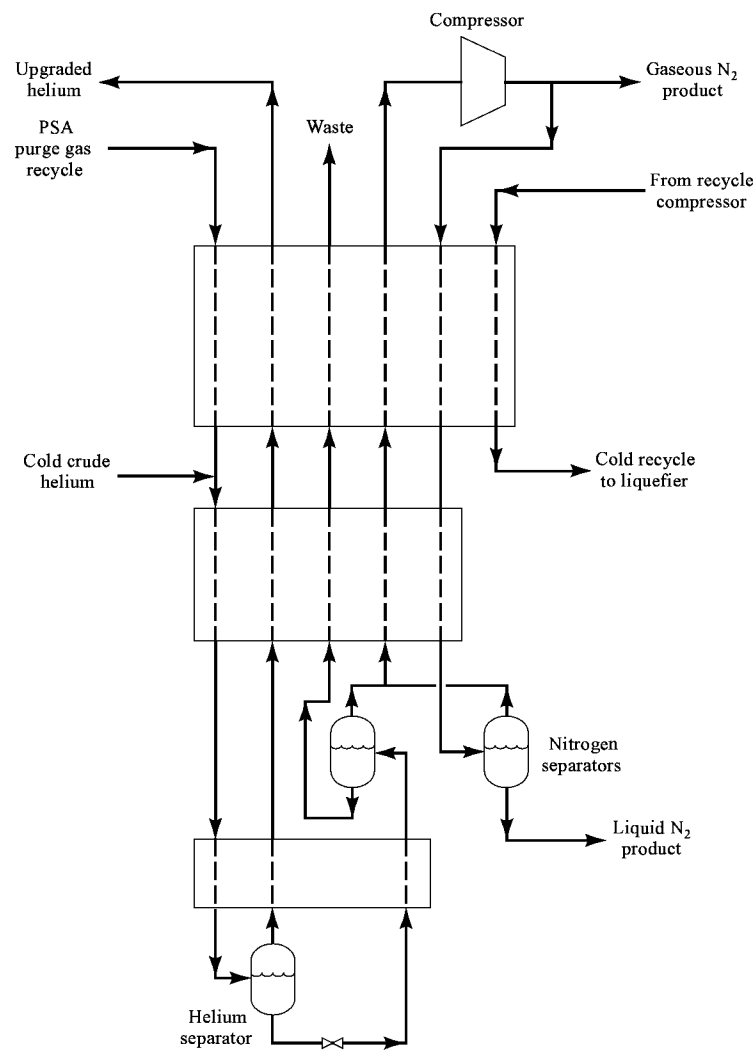


Fig. 10. Crude helium upgrading process.

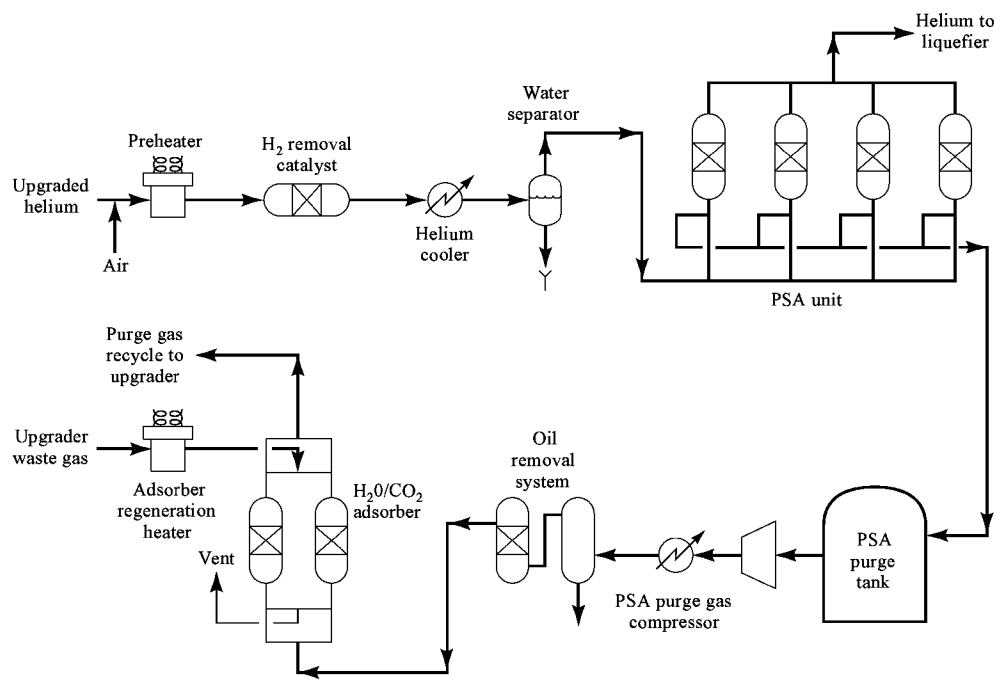


Fig. 11. Helium purification process.

Helium is liquefied by cooling to 80 K with either liquid nitrogen or the exhaust from a helium expander, then cooled below 80 K with refrigeration supplied by several helium expanders, followed by expansion through a dense-fluid wet expander. The number of expanders used in a liquefaction cycle varies with the plant capacity. Figure 12 shows a liquefier with 5 expanders (28). Helium is usually cooled at 2 MPa pressure, with the main flow from the exhaust of the expanders returning to compression at 0.2 MPa (29 psi), and a small stream from liquid Dewar boil-off and liquid trailer returning at 0.1 MPa. Heat exchangers in the liquefier are brazed aluminum plate-fin type, designed for temperature approaches as tight as 0.1 degree K to minimize energy requirements. The helium expanders are usually centrifugal with either oil- or gas-bearings, and are braked by either an oil cup loader or a closed-loop helium compressor. The main recycle compressors are reciprocating or (more typically) oil-flooded screw machines.

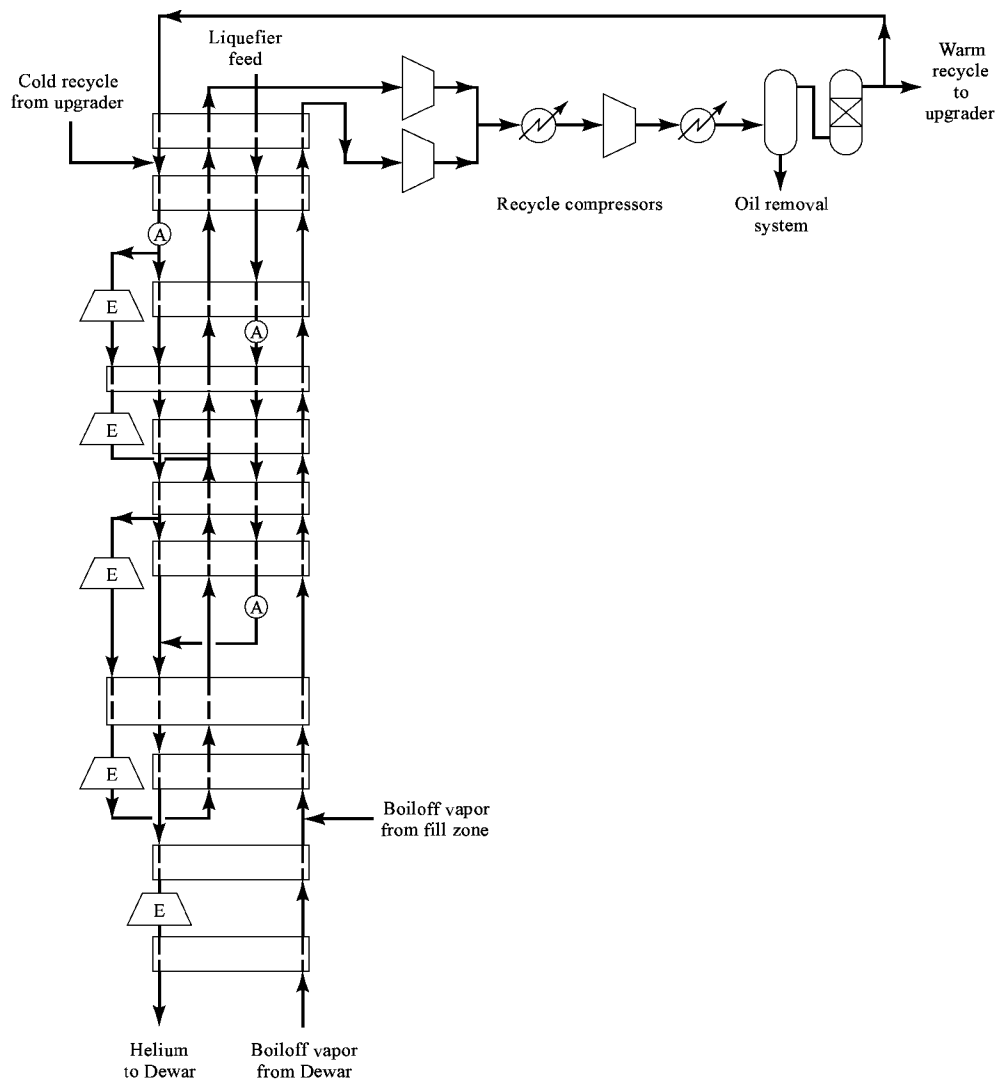


Fig. 12. Helium liquefaction process. A represents adsorber; E, expander.

Large helium refrigerators are used to cool superconducting magnets to 4.5 K in accelerators for high-energy physics experiments (29). The process configuration for a helium refrigerator is similar to that for a liquefier, and the refrigerator may produce some liquid. The process flows will vary with the ratio of 4 K heat load relative to the liquid production rate, because a refrigerator has balanced flows cooling and warming, with the heat load at the cold end, whereas a liquefier heat load is distributed over the temperature range from 300 to 4 K. Temperatures below 4.2 K are achieved by subatmospheric pressures maintained by cold compressors or by warm vacuum pumps. A refrigerator liquefier can trade 100 W of refrigeration at 4 K for 1 g s of liquid helium product. Liquid helium refrigerators are being considered for superconducting generators, transmission lines, and electric storage devices.

Equipment

Machinery. Compressor selection for a cryogenic process plant depends on the fluid, the volumetric flowrate, the pressures involved, the compressor efficiency, and the cost of energy and capital. Centrifugal compressors are lower in installed cost than reciprocating machines, and are preferred if the volumetric flows and pressures of the process allow them to be applied. For large volumetric flows, axial compressors are used. At very high pressures with small volumetric flows, reciprocating machines are required. Sometimes a combination of more than one kind of compression stage is used to yield the most cost effective and efficient system. For example, an oxygen compression system delivering oxygen at about 10 MPa (1450 psi) might employ several centrifugal stages followed by several reciprocating stages.

Gases of low molecular weight are difficult to compress in centrifugal compressors because of the low pressure rise imparted by a wheel that generates a reasonable head rise. This results in a large number of wheels being required to achieve a reasonable total pressure rise. Consequently, hydrogen and helium are seldom compressed using centrifugal compressors. Oil-flooded screw compressors are commonly used for helium processes operating at pressures under 2.4 MPa (348 psi), which is typical for helium liquefiers and refrigerators. Oil removal equipment must be used following lubricated reciprocating compressors or oil-flooded screw compressors.

Axial compressors have been applied to baseload LNG plants, but currently the majority of those plants use centrifugal compressors. Drivers for cryogenic plant compressors can be electric motors, internal combustion gas engines, gas turbines, or steam turbines. Cold compressors are in use with suction temperatures around 4 K to provide temperatures below 4 K by reducing the pressure under which helium is boiling. These cold compressors use magnetic bearings.

Expanders provide refrigeration by extracting work from a fluid, thereby reducing its temperature. Gas expanders can be either reciprocating or centrifugal, and can be loaded (braked) by electric generators, gas blowers or compressors, oil-film cups, or oil pump brakes. The extracted work can be usefully recovered in an electric generator or compressor. Centrifugal expanders are lower in first cost and in maintenance cost. Reciprocating expanders may be required for low volumetric flows and or high expansion pressure ratios of low molecular weight gases. Bearings for gas expanders can be oil or gas bearing type (static or dynamic). Reverse-running liquid pumps have been used as liquid expanders. Isentropic efficiencies for gas expanders can be as high as 85–90% for machines with high discharge volumetric flow.

Heat Exchangers. The two most prominent types of heat exchangers used in cryogenic service are the coil wound, tube-in-shell exchanger and the plate and fin (core) exchanger (30) (see HEAT EXCHANGE TECHNOLOGY). Since cryogenic process efficiency is highly dependent on maintaining close temperature approaches between all the cooling and warming streams, these heat exchangers are normally designed to accommodate at least three, and as many as eight or nine, countercurrent streams in a single unit. Accurate thermodynamic and physical property data as well as accurate heat transfer and pressure drop correlations are required for the design of these types of heat exchangers. The design must provide for uniform flow distribution to achieve close temperature approaches and must account for longitudinal heat conduction.

The coil wound heat exchanger consists of multiple tubes helically wound on a mandrel, usually with spacers between each tube layer. The tubes are inserted into tube sheets at both ends of the tube bundle, with separate tube sheets to accommodate each tube circuit. The tube bundle is enclosed in a shell with inlet and outlet nozzles for the shellside fluid. This type of heat exchanger is usually constructed of aluminum or stainless steel. Large aluminum

coil-wound heat exchangers for base-load LNG plants may be up to 5 m in diameter, 45 m long, and weigh up to 200 metric tons.

The plate and fin heat exchanger consists of corrugated sheets (fins) stacked between flat sheets (plates). The corrugated sheets are usually perforated or serrated to enhance the heat transfer performance of the exchanger. The stack is brazed in a closely controlled vacuum furnace and headers and nozzles are then welded to the brazed stack to form the separate circuits of the unit. This type of heat exchanger is normally constructed of aluminum, but stainless steel units are also becoming available.

Distillation Columns. In a cryogenic air separation plant, distillation (qv) accounts for the major fraction of the total energy consumption. The low relative volatility characteristic of many cryogenic separations requires the use of many stages. Columns operating at low pressures are consequently designed for a low pressure drop per theoretical stage of separation. Sieve trays with small perforations (~1 mm) provide excellent efficiency at low tray spacing for the clean fluids of cryogenic distillation processes. The minimum pressure drop across distillation sieve trays is set by the requirement to maintain a stable biphasic zone of gas bubbling through liquid, corresponding to about 35 mm of liquid per theoretical stage. A double column of the type shown in Figure 1 may use as many as 150 trays. Therefore it is important to keep the spacing between the trays as small as possible to minimize capital cost. Usually tray spacing is in the 10 to 20 cm range. Important manufacturers each have their own design of the tray geometry, including multipass cross flow trays, split cross flow trays, and circular flow trays. Recently, low pressure drop structured packings have been used to replace sieve trays in some sections of the upper column and the argon column of air-separation units.

In some cryogenic hydrocarbon separation plants such as a nitrogen rejection unit, both flowrate and feed composition may change over the life of the plant. This requires the use of valve or bubble cap trays with high turndown capability for distillation. In some applications, the presence of foaming or high froth liquids requires tray spacing as high as 60 cm.

Insulation. Cryogenic insulation should economically reduce heat leak into the system so that its impact on the overall refrigeration requirement is minimized. Insulation can be categorized as unevacuated bulk type (eg, purged rockwool or perlite), rigid foam (eg, foam-glass or urethane), vacuum-jacketed (VJ), evacuated powder (eg, perlite), and multilayer insulation (MLI) (eg, evacuated aluminized Mylar) (see INSULATION, THERMAL).

Process equipment for cryogenic plants is often enclosed in a cold box consisting of a steel frame with panels that is filled with perlite or rockwool and purged with a gas that will not condense on the exterior of the equipment and piping. This method of insulation is commonly used on plants for air separation, hydrocarbon recovery, hydrogen purification, and nitrogen refrigeration. Equipment supports or suspension systems are designed for low conduction heat leak using materials with low thermal conductivity, long conduction paths, and small cross-sections. Small refrigerators commonly use evacuated MLI. For the large equipment sizes of baseload LNG plants, individual equipment items and piping are separately insulated with urethane. Lines to transfer liquid cryogenics can be insulated with rigid foam, vacuum, evacuated MLI, or evacuated MLI with heat shield, depending on the impact of heat leak and the frequency of cooldown. Bayonet fittings and valves for VJ lines are designed for low heat leak using long conduction paths with small cross sections.

Storage tanks for large volumes (>1000 m³) of LNG, liquid nitrogen and liquid oxygen often use unevacuated perlite insulation that is pressurized with boil-off gas from the tank's liquid. Tanks for similar volumes of liquid H₂ are usually insulated with evacuated perlite. Smaller tanks for LNG, liquid nitrogen, and liquid oxygen use evacuated perlite insulation. Tanks for liquid helium and smaller ones for liquid hydrogen are insulated with evacuated MLI, and the liquid helium tanks usually incorporate a heat shield using boiloff vapor or liquid nitrogen to intercept part of the heat leak.

Transport vehicles for cryogenic liquids have an inner vessel holding the liquid and a surrounding vacuum perlite or MLI insulation, all enclosed within an outer vessel. Liquid helium vehicles are insulated between the inner and outer vessels by vacuum MLI and a heat shield cooled by liquid nitrogen which is carried on board. These liquid helium containers can be filled with liquid helium in the U.S. and transported by ship to Europe or Asia without venting helium for 45 days (31). Barges and rail-cars to transport cryogenics are in everyday service in the U.S. in support of the space program. LNG ships of 125,000 m³ cargo capacity use rigid foam insulation.

Safety

The possibility of an uncontrolled release of a cryogenic fluid such as liquid oxygen, liquid methane or liquid hydrogen from storage and during handling must be carefully considered during the design of a cryogenic facility. The level of risk is often reduced by providing dikes for secondary containment of liquid spills. Procedures to protect personnel from cryogenic burns and asphyxia and to protect the nearby equipment from embrittlement failure must be carefully considered and followed. When trace impurities in the feed streams can lead to the combination of an oxidant with a flammable cryogen (eg, solid oxygen in liquid hydrogen) or a combustible with an oxidant (eg, acetylene in liquid oxygen), special precautions must be taken to eliminate them. Many materials react with pure oxygen, so care must be taken in the selection of materials in contact with oxygen and in the cleaning of oxygen systems prior to use. Potential ignition sources must be minimized, particularly in oxygen compression and in systems for handling oxygen at elevated pressures (32).

BIBLIOGRAPHY

"Cryogenics" in *ECT* 1st ed., Suppl. 2, pp. 272–281, by H. R. Morrison, Linde Company, Division of Union Carbide Corporation; in *ECT* 2nd ed., Vol. 6, pp. 471–481, by H. R. Morrison, Linde Company, Division of Union Carbide Corporation; in *ECT* 3rd ed., Vol. 7, pp. 227–242, by C. L. Newton, Air Products and Chemicals, Inc.

1. B. A. Hands, ed., *Cryogenic Engineering*, Academic Press, New York, 1986.
2. W. H. Isalski, *Separation of Gases*, Clarendon Press, Oxford, 1989.
3. K. D. Timmerhaus and T. M. Flynn, *Cryogenic Process Engineering*, Plenum Press, New York, 1989.
4. R. A. Thiévenot, *History of Refrigeration Throughout the World*, trans. from French by J. C. Fidler, International Institute of Refrigeration, Paris, 1979.
5. R. E. Latimer, *Chem. Eng. Prog.* **63**(2), 35 (1967).
6. R. Agrawal and D. W. Woodward, "Impact of Structured Packing on Distillation of Argon from Air", *Proceedings of the XVIIIth International Congress of Refrigeration*, Vol. I, Montréal, Canada, 1991, p. 162.
7. **U.S. Pat. 5,049,173** (Sept. 17, 1991), T. E. Cormier, R. Agrawal, D. W. Woodward and A. Prentice (to Air Products and Chemicals, Inc.).
8. **U.S. Pat. 3,817,046** (June 18, 1974), I. Aoki and Y. Kitsukawa (to Chinzoda Chemical Engineering and Construction Co., Ltd., and Air Products and Chemicals, Inc.).
9. **U.S. Pat. 4,339,253** (July 13, 1982), E. Caetani and H. Paradowski (to Compagnie Française d'Etudes and de Construction "Technip").
10. **U.S. Pat. 4,525,185** (June 25, 1985), C. L. Newton (to Air Products and Chemicals, Inc.).
11. **U.S. Pat. 4,545,795** (Oct. 8, 1985), Y. N. Liu and J. W. Pervier (to Air Products and Chemicals, Inc.).
12. H. C. Rowles, D. M. Nicholas, H. L. Vines and M. F. Hilton, *Energy Prog.* **6**(1), 25 (1986).
13. W. K. Lam, W. J. Stupin and G. Christensen, "Recover Valuable Off-Gases by the Braun ROE Process," paper presented at *AIChE National Meeting*, New Orleans, La., Apr. 6–10, 1986.
14. C. L. Newton, *Chem. and Process Eng.*, 51, (Dec. 1967).
15. C. L. Newton, *Cryogenic Eng. News*, 50 (Aug., 1967); 24 (Sept., 1967).
16. W. J. Moore, *Physical Chemistry*, Prentice-Hall, Englewood Cliffs, N.J., 1955.
17. R. B. Scott, *Cryogenic Engineering*, Met-Chem Research Inc., Boulder, Colo., 1988.
18. M. S. Lipman, H. Chueng and O. P. Roberts, *Chem. Eng. Prog.* **59**(8), 49 (1963).
19. R. O. Voth, J. D. Siegworth, J. F. Welch and M. J. Hiza, *Process Design Criteria, Data and Metrology Related to Slush Hydrogen Technology*, NASA Tech. Memo 1006, NASA Langley Res. Cntr., Mar., 1987.
20. G. A. Lucadamo, D. P. Bernhard and H. C. Rowles, *Gas Separation and Purification* **1**, 94 (1987).
21. M. G. Brahn and R. H. McCue, "Advanced Recovery System for Refinery Petrochemical Applications", paper presented at *AIChE National Meeting*,

Orlando, Fla., Mar. 21, 1990.

22. H. L. Vines, *Chem. Eng. Prog.*, 46 (Nov. 1986).

23. C. A. Pruitt and J. V. O'Brien, *Oil and Gas J.*, 78 (Oct. 9, 1989).

24. L. S. Gaumer, *Chem. Eng. Prog* **63**(5), 72 (1967).

25. A. Goethe and D. J. Mawer, "A New Integrated Nitrogen Rejection Process with NGL Recovery," *paper presented at Gas Processing Association (GPA) Annual Convention*, Denver, Colo., Mar. 16–18, 1987.

26. R. D. McCarty, *Thermophysical Properties for Helium-4 from 2 to 1500K with Pressures to 1000 Atmospheres*, NBS-TN-631, U.S. Dept. of Commerce, Nov. 1972.

27. S. W. Van Sciver, *Helium Cryogenics*, Plenum Press, New York, 1986.

28. R. R. Olsen and R. F. Pahade, in M. McAshan, ed., *Supercollider I*, Proc. of IISSC Conf., Feb. 8–10, 1989, New Orleans, Plenum Press, New York, 1989.

29. J. R. Sanford and D. M. Matthews, eds., *Site-Specific Conceptual Design of the Super-conducting Super Collider, SSCL-SR-1056*, SSC Laboratory, Dallas, Tex., July, 1990.

30. R. E. Lowe, *Chem. Eng.* **17**, 131 (Aug., 1987).

31. Technical literature, Gardner Cryogenics, Bethlehem, Penn., 1992.

32. B. R. Dunbobbin, J. G. Hansel and B. L. Werley, in J. M. Stoltzfus and K. McIlroy, eds., *Flammability and Sensitivity of Materials in Oxygen Enriched Atmospheres*, Vol. 5, ASTM STP 1111, American Society for Testing and Materials, Philadelphia, 1991.

Rakesh Agrawal
Howard C. Rowles
Glenn E. Kinard
Air Products and Chemicals, Inc.



CRYSTALLIZATION

Crystallization is one of the oldest unit operations in the portfolio of industrial and or laboratory separations. Almost all separation techniques involve formation of a second phase from a feed, and processing conditions must be selected that allow relatively easy segregation of the two or more resulting phases. This is a requirement for crystallization also, and there are a variety of other properties of the solid product that must be considered in the design and operation of a crystallizer. Interactions among process, function, product, and phenomena important in crystallization are illustrated in Figure 1.

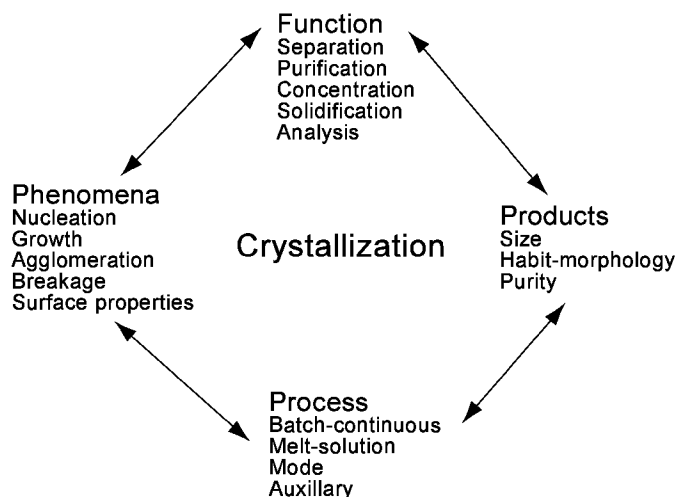


Fig. 1. Crystallization.

Function. Figure 1 lists several possible functions that can be achieved by crystallization: separation, purification, concentration, solidification, and analysis. A few examples follow.

Separation. Sodium carbonate (soda ash) is recovered from a brine by first contacting the brine with carbon dioxide to form sodium bicarbonate. Sodium bicarbonate has a lower solubility than sodium carbonate, and it can be readily crystallized. The primary function of crystallization in this process is separation; a high percentage of sodium bicarbonate is solidified in a form that makes subsequent separation of the crystals from the mother liquor economical. With the available pressure drop across filters that separate liquid and solid, the capacity of the process is determined by the rate at which liquor flows through the filter cake. That rate is set by the crystal size distribution produced in the crystallizer.

Separation of a chemical species from a mixture of similar compounds can also be achieved by melt crystallization, which is, for example, an important means of separating *para*-xylene (*p*-xylene) from the ortho and meta isomers. *p*-Xylene is crystallized at the top of a vertical column and crystals are moved downward countercurrently to liquid. The liquid flowing upward is generated by adding heat to melt the crystals at the bottom of the column; a portion of the melt is removed as product and the remainder flows up the column to contact the downward-flowing crystals. Effluent mother liquor, consisting almost entirely of the ortho and meta isomers of xylene, is removed from the top of the column.

Concentration. The concentration of fruit juice requires removal of solvent (water) from the natural juice. This is commonly done by evaporation, but the derived juices may lose flavor components or undergo thermal degradation during evaporation. In freeze concentration, solvent is crystallized (frozen) in a relatively pure form to leave behind a solution with a solute concentration higher than the original mixture. Significant advantages in product taste have been observed in the application of this process to concentration of certain fruit juices.

Purification. The objective of crystallization also can be purification of a chemical species. For example, L-isoleucine (an essential amino acid) is separated by crystallization from a fermentation broth that has been filtered and subjected to ion exchange. The recovered crystals contain impurities deleterious to use of the product, and these crystals are, therefore, redissolved and recrystallized to enhance purity.

Solidification. Production of a product in a form suitable for use and acceptable to the consumer also may be an objective of a crystallization process. For example, the appearance of sucrose (sugar) varies with local customs, and deviations from that custom could lead to an unacceptable product. A final crystallization may thus be called for to bring the product appearance into compliance with expectations.

Analysis. Many analytical procedures calling for determination of molecular structure are aided by crystallization or require that the unknown compound be crystalline. Methodologies coupling crystallization and analytical procedures will not be covered here (see X-RAY TECHNOLOGY)

Products. In all of the instances in which crystallization is used to carry out a specific function, product requirements are a central component in determining the ultimate success of the process. These requirements grow out of how the product is to be used and the processing steps between crystallization and recovery of the final product. Key determinants of product quality are the size distribution (including mean and spread), the morphology (including habit or shape and form), and purity. Of these, only the last is important with other separation processes.

Crystal size distribution (CSD) determines several important processing and product properties, including crystal appearance, separation of crystals from liquor, reactions, dissolution and other processes and properties involving surface area of the crystalline product, crystal transportation, and crystal storage. In fact, experience indicates that a large fraction of crystallizer troubleshooting cases have been initiated to solve problems associated with inadequate throughput of filters or centrifuges; when solutions are found they generally involve manipulation of CSD.

It is often important to control the CSD of pharmaceutical compounds, eg, in the synthesis of human insulin, which is made by recombinant DNA techniques (1). The most favored size distribution is one that is monodisperse, ie, all crystals are of the same size, so that the rate at which the crystals dissolve and are taken up by the body is known and reproducible. Such uniformity can be achieved by screening or otherwise separating the desired size from a broader distribution or by devising a crystallization process that will produce insulin in the desired form. The latter of these options is preferable, and considerable effort has been expended in that regard.

Process. In each of the systems discussed above there is a need to form crystals, to cause the crystals to grow, and to separate the crystals from residual liquid. There are various ways to accomplish these objectives leading to a multitude of processes that are designed to meet requirements of product yield, purity, and, uniquely, crystal size distribution.

Phenomena. The critical phenomena in crystallization are, as shown in Figure 1, nucleation and growth kinetics, interfacial phenomena, breakage, and agglomeration. Nucleation leads to the formation of crystals, either from a solution or a melt. Growth is the enlargement of crystals caused by deposition of solid material on an existing surface. The relative rates at which nucleation and growth occur determine the crystal size distribution; qualitatively, when the rate of nucleation is high relative to growth rate, crystals formed are small and numerous. Agglomeration is the formation of a larger

particle through two or more smaller particles (crystals) sticking together. It is prevalent in many processes, and agglomeration can be essential for solid–liquid separation or it can be undesirable because it may adversely affect crystal quality. Breakage of crystals is almost always undesirable because it is detrimental to crystal appearance and it can lead to excessive fines and have a deleterious effect on crystal purity. Interfacial phenomena influence solid–liquid separation, flow characteristics of slurries, agglomeration, and crystal morphology.

Solid–Liquid Equilibria and Mass and Energy Balances

Solubility. Solid–liquid equilibrium, or the solubility of a chemical compound in a solvent, refers to the amount of solute that can be dissolved at constant temperature, pressure, and system composition; in other words, the maximum concentration of the solute in the solvent at static conditions. In a system consisting of a solute and a solvent, specifying system temperature and pressure fixes all other intensive variables. In particular, the composition of each of the two phases is fixed, and solubility diagrams of the type shown for a hypothetical mixture of R and S in Figure 2 can be constructed. Such a system is said to form an eutectic, ie, there is a condition at which both R and S crystallize into a solid phase at a fixed ratio that is identical to their ratio in solution. Consequently, there is no change in the composition of residual liquor as a result of crystallization.

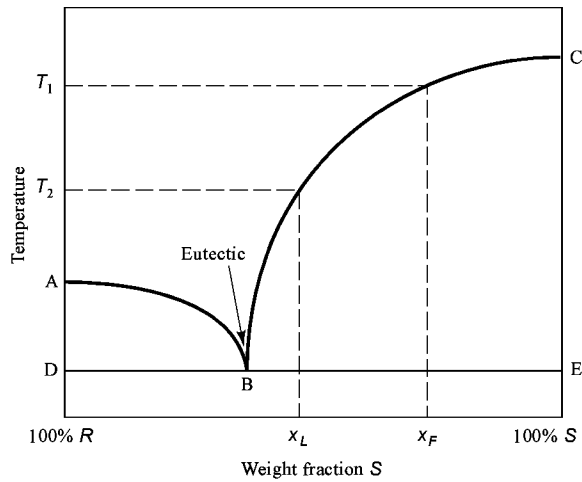


Fig. 2. Solubility diagram for a hypothetical system. The curves AB and BC represent solution compositions that are in equilibrium with solids whose compositions are given by the lines AD and CE. If AD and CE are vertical along the respective axes, the crystals are pure R and S, respectively. Crystallization from any solution whose composition is to the left of the vertical line through point B produces crystals of pure R, whereas solutions to the right of the line produce crystals of pure S. A solution whose composition falls on the line through B produces a solid mixture that has a composition identical to the liquid solution.

Several features of the hypothetical system in Figure 2 can be used to illustrate proper selection of crystallizer operating conditions and limitations placed on the operation by system properties. Suppose a saturated solution at temperature T_1 is fed to a crystallizer operating at temperature T_2 . Because the feed is saturated, the weight fraction of S in the feed is given as shown in Figure 2. The maximum crystal production rate $P_{\max}$ from such a process depends on the value of T_2 and is given by

$$P_{\max} = F x_F - L x_L \tag{1}$$

where F is the feed rate to the crystallizer, and L is the solution flow rate leaving the crystallizer. No other stream is fed to or removed from the crystallizer. Note that the lower limit on T_2 is given by the eutectic point B.

Figure 3 presents the equilibrium behavior of magnesium sulfate in water, and it is illustrative of systems that form hydrated salts. Equilibrium solution concentrations are plotted as curves ab , bc , cd , de , and ef ; the solid phases that are in equilibrium with these solutions have compositions given by the lines ag , bi , jk , lm , and no , respectively. Ice is the solid phase whose composition is given by ag , and crystals containing differing ratios of water of hydration to magnesium sulfate constitute the solids represented by the other lines. Specifically, the line no represents magnesium sulfate monohydrate ($\text{MgSO}_4\cdot\text{H}_2\text{O}$), which has one water molecule per molecule of magnesium sulfate, whereas the lines ml , kj , and ib represent the hexahydrate, heptahydrate, and dodecahydrate forms, respectively. The weight fraction of MgSO_4 in each of the crystal forms is shown in Figure 3, and as with all crystalline materials having water of hydration, the solute balance of equation 1 must be modified to read

$$x_c P_{\max} = F x_F - L x_L \tag{2}$$

where x_c is the mass fraction of solute in the crystal, eg, x_c is 0.488 when the crystalline substance is magnesium sulfate heptahydrate. Differences in the forms of magnesium sulfate crystals affect the dependence of solubility on temperature, which is reflected by the slopes of the solution composition curves.

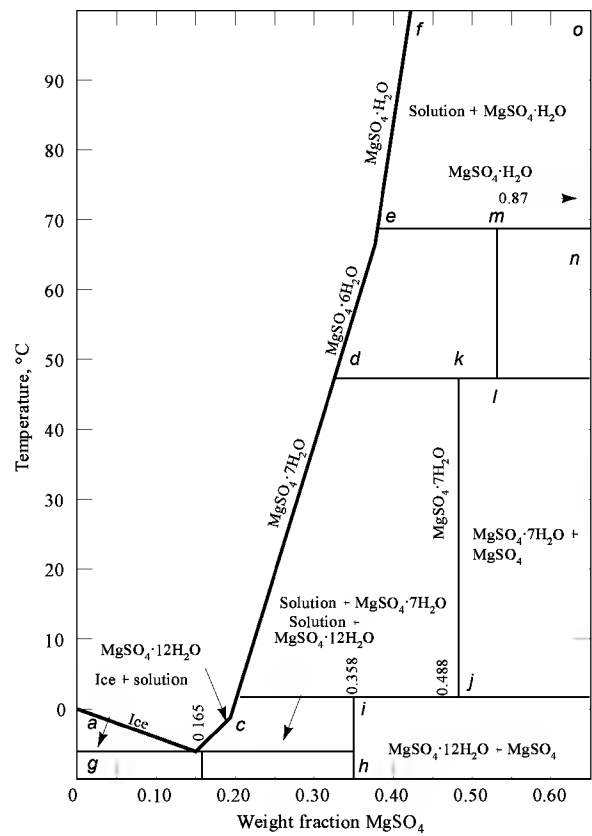


Fig. 3. Solubility diagram for magnesium sulfate in water.

The dependence of solubility on temperature affects the mode of crystallization. For example, Figure 4 shows that the solubility of potassium nitrate is strongly influenced by the system temperature but that temperature has little influence on the solubility of sodium chloride. As a consequence, a reasonable yield of KNO₃ crystals can be obtained by cooling a saturated feed solution; on the other hand, cooling a saturated sodium chloride solution accomplishes little crystallization, and evaporation is required to increase the yield of sodium chloride crystals.

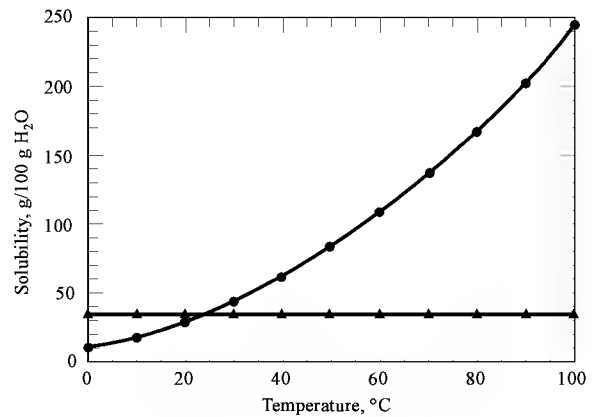


Fig. 4. Solubilities of NaCl ▲, and KNO₃ ●, in water.

The production of many high value chemicals requires maximizing separation from a relatively dilute solution. It is common in such instances to use a combination of methods to reduce solute solubility in the feed solution. Figure 5, for example, illustrates that the addition of methanol to a saturated aqueous solution of L-serine can reduce solubility by more than an order of magnitude.

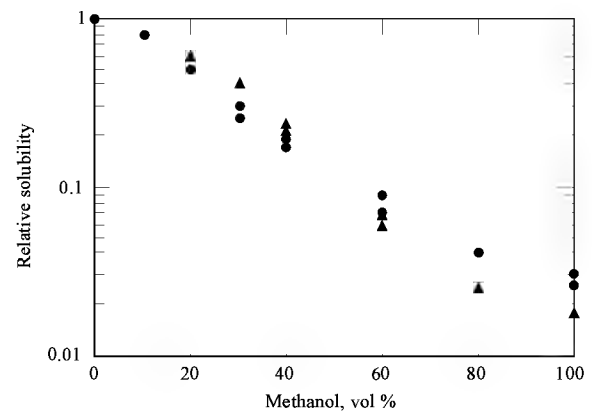


Fig. 5. Solubilities of L-serine in mixtures of methanol and water. Relative solubility — solubility in CH₃OH-H₂O : solubility in H₂O; ●, 10°C; ▲, 30°C. Reproduced by permission of the American Institute of Chemical Engineers (4).

Solubility data can be found in a variety of units, and conversion from one set of units to another often is required before computation of yield can be performed. Guides to such conversions are available. It is often most convenient, however, to express solubility and compositions in mixed streams in terms of mass ratios, ie, mass of solute per mass of solvent.

Supersaturation. The thermodynamic driving force for both crystal nucleation and growth is the key variable in setting the mechanisms and rates by which these processes occur. Supersaturation is defined rigorously as the deviation from thermodynamic equilibrium, which is the difference between the chemical potential of the solute at the existing conditions of the system μ and the chemical potential of the solute equilibrated at the system conditions μ^* . Less abstract definitions involving measurable system properties such as temperature, concentration, or mass or mole fraction also have been used to express supersaturation. Consider, for example, a system at temperature T with a solute concentration c . A saturated solution having a concentration c would be at temperature T^* , whereas a saturated solution having a temperature T would have a solute concentration c^* .

Expressions of supersaturation can then be formulated as follows: (1) the difference between the chemical potential of the system and the chemical potential of at saturation, $\mu - \mu^*$, where the chemical potential is a function of both temperature and concentration; (2) the difference between the solute concentration and the concentration at equilibrium, $c - c^*$; (3) the difference between the system temperature at equilibrium and the actual temperature, $T^* - T$; (4) the ratio of the solute concentration and the equilibrium concentration, c / c^* , which is known as relative saturation; or (5) the ratio of the difference between the solute concentration and the equilibrium concentration to the equilibrium concentration, $s = (c - c^*) / c^*$, which is known as relative supersaturation. This term has often been represented by σ ; s is used here because of the frequent use of σ for interfacial energy or surface tension and for variance in distribution functions.

Any of these definitions of supersaturation can be used over a moderate range of system conditions. For example, a difference in chemical potential $\Delta\mu = \mu - \mu^*$ is proportional to both $c - c^*$ and $T^* - T$ over a modest range of conditions. Of the five expressions given, however, the second is most useful in calculating the yield from a crystallizer, the third provides information that may be most useful in the control of a crystallizer, and the fifth is most commonly used in correlating the dependence of nucleation and growth kinetics on supersaturation.

Although the first of these expressions is the most nearly fundamentally correct, it is difficult to evaluate because of inadequate capabilities of determining chemical potential as a function of temperature and composition. The next most appropriate driving force for crystal growth is that given by the last of the possibilities, ie, $s = (c - c^*) / c^*$. This definition of supersaturation will be used throughout the ensuing discussion, but it should be recognized that the validity of doing so is limited to low supersaturations. Solute concentrations used in determining supersaturation should include any water of hydration associated with the solute in the crystal at equilibrium.

Mass and Energy Balances. The formulation of mass and energy balances follows procedures outlined in many basic texts (2). The use of solubilities to calculate crystal production rates from a cooling crystallizer was demonstrated by the discussion of equations 1 and 2. Subsequent to determining the yield, the rate at which heat must be removed from such a crystallizer can be calculated from an energy balance:

$$F \hat{H}_F - P \hat{H}_C + L \hat{H}_L + Q$$

(3)

where F , P , and L are feed rate, crystal production rate, and mother liquor flow rate, respectively; $\hat{H}$ is the specific enthalpy of the stream corresponding to the subscript; and Q is the required rate of heat transfer. As F , P , and L are known or can be calculated from a simple mass balance, determination of Q requires methods of estimating specific enthalpies.

If appropriate enthalpy data are unavailable, estimates can be obtained by first defining reference states for both solute and solvent. Often the most convenient reference states are crystalline solute and pure solvent at an arbitrarily chosen reference temperature. The reference temperature selected usually corresponds to that at which the heat of crystallization $\Delta\hat{H}_C$ of the solute is known. The heat of crystallization is approximately equal to the negative of the heat of solution. For example, if the heat of crystallization is known at T_{ref} , then reasonable reference conditions would be the solute as a solid and the solvent as a liquid, both at T_{ref} . The specific enthalpies then could be evaluated as

$$\hat{H}_F = x_F \Delta\hat{H}_C + C_{pF}(T - T_{\text{ref}})$$

(4)

$$\hat{H}_C = C_{pC}(T - T_{\text{ref}})$$

(5)

$$\hat{H}_L = x_L \Delta\hat{H}_C + C_{pL}(T - T_{\text{ref}})$$

(6)

where x_F and x_L are the mass fractions of solute in the feed and mother liquor, respectively. All that is required now to determine the required rate of heat transfer is the indicated heat capacities, which can be estimated or measured experimentally.

Now suppose some of the solvent is evaporated in the crystallizer, as is shown in Figure 6. Independent balances can be written on total mass and solute:

$$F = V + L + P$$

(7)

$$x_F F = x_L L + x_C P$$

(8)

where F = feed rate, V = vapor withdrawal rate, L = liquid (filtrate) withdrawal rate, and P = crystal production rate, and x_j = solute content of stream j in units consistent with flow rate. There is an equilibrium expression relating x_L , x_C and the temperature or pressure at which the operation is conducted. In addition, an energy balance must be satisfied:

$$F \hat{H}_F + Q = V \hat{H}_V + L \hat{H}_L + P \hat{H}_C$$

(9)

The specific enthalpies ($\hat{H}$) in equation 9 can be determined as described earlier, provided the temperatures of the product streams are known. Evaporative cooling crystallizers operate at reduced pressure and may be considered adiabatic ($Q = 0$). As with of many problems involving equilibrium relationships and mass and energy balances, trial-and-error computations are often involved in solving equations 7 through 9.

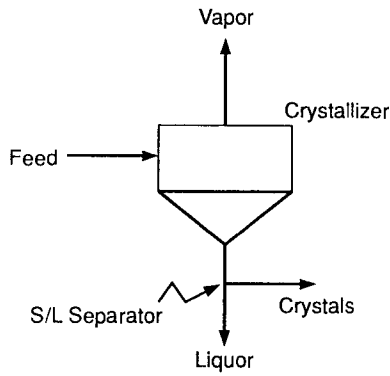


Fig. 6. Schematic diagram of a simple crystallizer.

The mass balance on a crystallizer is related to the growth kinetics that occur within the unit. This may be illustrated by considering systems in which crystal growth kinetics are sufficiently fast to use essentially all of the supersaturation provided by the crystallizer. Under such conditions (referred to in the crystallization literature as class II or fast-growth behavior), the solute concentration in the mother liquor can be assigned a value corresponding to saturation. Alternatively, should supersaturation in the mother liquor be so great as to affect the solute balance, the operation is said to follow class I or slow-growth behavior. An expression coupling the rate of growth to a solute balance must be used to describe such a system.

Crystallization Kinetics

Along with operating variables of the crystallizer, nucleation and growth determine such crystal characteristics as size distribution, purity, and shape or habit.

Nucleation. Crystal nucleation is the formation of an ordered solid phase from a liquid or amorphous phase. Nucleation sets the character of the crystallization process, and it is, therefore, the most critical component in relating crystallizer design and operation to crystal size distributions.

Mechanisms. Classical nucleation theory is based on homogeneous and heterogeneous mechanisms, both of which call for the formation of crystals through a process of sequentially combining the constituent units that form a crystal. Heterogeneous and homogeneous mechanisms are referred to as primary nucleation because existing crystals play no role in the nucleation. Primary nucleation can be illustrated by considering a hypothetical experiment in the context of the solubility data in Figure 7. Assume that a solution is at a concentration and temperature corresponding to point A on the figure. The solution is undersaturated, so any crystals present in the system would dissolve. If the concentration is increased at constant temperature, for example by evaporation, the path followed would cause the solution to reach saturated conditions at point B. Once the concentration becomes greater than that at B, the solution is supersaturated and any crystals present in the system would grow. However, experience shows that nucleation would not occur until the concentration reaches point C, which defines what is called the metastable limit. If this procedure is repeated at various temperatures, a metastable limit curve could be drawn as shown.

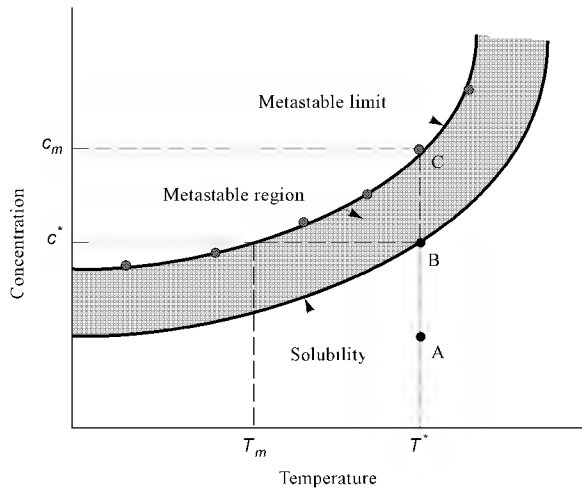


Fig. 7. Solubility and primary nucleation in a hypothetical experiment.

Within the metastable region, primary nucleation will not occur, and the width of the metastable zone, as reflected by $c_m - c^*$ or $T^* - T_m$, varies from one substance to another. Furthermore, the width can vary for the same substance with composition, the presence of impurities that alter interfacial tension, and various other factors.

Both homogeneous and heterogeneous mechanisms require relatively high supersaturation, and they exhibit a high order dependence on supersaturation. These factors often lead to production of excessive fines in systems where primary nucleation mechanisms are important. The classical theoretical treatment of primary nucleation results in the expression (5):

$$B^0 = A \exp \left(\frac{16\pi\sigma^3 v^2}{3k^3 T^3 [\ln(s + 1)]^2} \right) \tag{10}$$

where k is the Boltzmann constant, σ is surface energy per unit area, v is molar volume, and A is a constant. This equation can be simplified by recognizing that $\ln(s + 1)$ approaches s as s approaches 0. So for small supersaturations,

$$B^0 = A \exp \left(\frac{16\pi\sigma^3 v^2}{3k^3 T^3 s^2} \right) \tag{11}$$

The most important variables affecting nucleation rate are shown by equations 10 and 11 to be interfacial energy, temperature, and supersaturation. The high order dependence of nucleation rate on supersaturation is especially important because a small swing in supersaturation may produce an enormous change in nucleation rate. This gives rise to the often-observed phenomenon of having a clear liquor transformed to a slurry of fine crystals with only a slight increase in supersaturation, for example, by decreasing the solution temperature.

The catalytic effect of solid particles (as in heterogeneous nucleation) is to reduce the energy barrier to formation of a new phase. This, in effect, can reduce the interfacial energy σ significantly.

The metastable limit can provide an empirical approach to modeling primary nucleation. This limit, which was first observed in 1951 (6), must be determined through experimentation, and nucleation rate is correlated with the following equation

$$B^0 = k(c - c_m)^i \tag{12}$$

where the equilibrium concentration c^* is less than the concentration at the metastable limit, c_m . Values of c_m are often very close to c^* for many inorganic systems, and satisfactory correlations have been obtained with c^* substituted for c_m in equation 12 (7); in other words,

$$B^0 = k(c - c^*)^i \tag{13}$$

where the parameters k and i must be evaluated from experimental data.

Secondary nucleation is crystal formation through a mechanism involving the solute crystals; crystals of the solute must be present for secondary nucleation to occur. Thorough reviews have been given (8,9).

Several features of secondary nucleation make it more important than primary nucleation in industrial crystallizers. First, continuous crystallizers and seeded batch crystallizers have crystals in the magma that can participate in secondary nucleation mechanisms. Second, the requirements for the mechanisms of secondary nucleation to be operative are fulfilled easily in most industrial crystallizers. Finally, low supersaturation can support secondary nucleation but not primary nucleation, and most crystallizers are operated in a low supersaturation regime that improves yield and enhances product purity and crystal morphology.

Secondary nucleation can occur as the result of several mechanisms that have been identified in selected systems and include the following. *Initial breeding* results from immersion of seed crystals in a supersaturated solution. It is thought to be caused by dislodging extremely small crystals that were formed on the surface of larger crystals during drying. Although this mechanism is unimportant in continuous and unseeded batch crystallization, it can be significant in the operation of seeded batch crystallizers. Several process variables have been shown to influence nucleation rates caused by initial breeding (10). *Contact nucleation* results from collisions of crystals with one another, and or crystallizer internals, and or an impeller in an agitator or circulation pump. The collision energy required for contact nucleation is small and does not result in macroscopic degradation (breakage) of the contacted crystal. *Shear breeding* results when supersaturated solution flows by a crystal surface and carries with it crystal precursors believed formed in the region of the growing crystal surface. In a study of nucleation of $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ by this mechanism, it was found (11) that high levels of supersaturation were required for it to produce significant numbers of nuclei.

Process Variables Affecting Contact Nucleation. Pioneering studies elucidated many factors affecting contact nucleation (12–14). The number of crystals produced by a controlled impact of an object with a seed crystal depends on energy of impact, supersaturation at impact, supersaturation at which crystals mature, hardness of the impacting object, area of impact, angle of impact, and system temperature. Although it is impossible to account quantitatively for all of these variables, certain generalizations can be drawn from the research on this nucleation mechanism.

Based on experimental observations, the following expression was proposed (15) for systems at constant supersaturation:

$$B^0 = k_N \exp(E - E_t) \tag{14}$$

The same researchers proposed that a relationship of impact energy E to crystallizer variables must include the mass of the impacting crystal m_c , the rotational velocity of the impeller providing mixing ω , and the fraction of the available energy actually transmitted to the crystal ϵ :

$$E = f(\omega, \epsilon, m_c) \tag{15}$$

Correlations of nucleation rates with crystallizer variables have been developed for a variety of systems. Although the correlations are empirical, a mechanistic hypothesis regarding nucleation can be helpful in selecting operating variables for inclusion in the model. Two examples are (1) the effect of slurry circulation rate on nucleation has been used to develop a correlation for nucleation rate based on the tip speed of the impeller (16) and (2) the scaleup of nucleation kinetics for sodium chloride crystallization provided an analysis of the role of mixing and mixer characteristics in contact nucleation (17). Published kinetic correlations have been reviewed through about 1979 (18). In a later section on population balances, simple power-law expressions are used to correlate nucleation rate data and describe the effect of nucleation on crystal size distribution.

Supersaturation has been observed to affect contact nucleation, but the mechanism by which this occurs is not clear. There are data (19) that infer a direct relationship between contact nucleation and crystal growth. This relationship has been explained by showing that the effect of supersaturation on contact nucleation must consider the reduction in interfacial supersaturation due to the resistance to diffusion or convective mass transfer (20).

Still another possible role of supersaturation is that it affects the solution structure and causes the formation of clusters of solute molecules. These clusters may participate in nucleation, although the mechanism by which this would occur is not clear. Evidence of the existence of cluster formation in supersaturated solutions has been presented for citric acid (21); while others have examined the phenomenon in greater detail (22,23).

The ease with which nuclei can be produced by contact nucleation is a clear indication that this mechanism is dominant in many industrial operations. Research on contact nucleation is continuing with the objective of building an understanding of the phenomenon that will allow its successful inclusion in models describing commercial systems.

Crystal Growth. At least two resistances determine growth kinetics, those associated with integration or incorporation of the crystalline unit (for example, solute molecules) into the crystal surface (lattice) and molecular or bulk transport of the unit from the surrounding solution to the crystal face. The primary concern here is with surface incorporation. A simple set of experiments in which the rate of advance of a crystal face is measured can be used to illustrate these two resistances. Data given in Figure 8 show the effect of solution velocity over a crystal face at three different conditions. As the solution velocity increases from low values, the growth rates also increase. At about 24 cm/s, however, the growth rates approach constant values. Such behavior indicates that both bulk mass transfer and surface incorporation are important below 24 cm/s but above this velocity, surface incorporation provides the dominant resistance to growth.

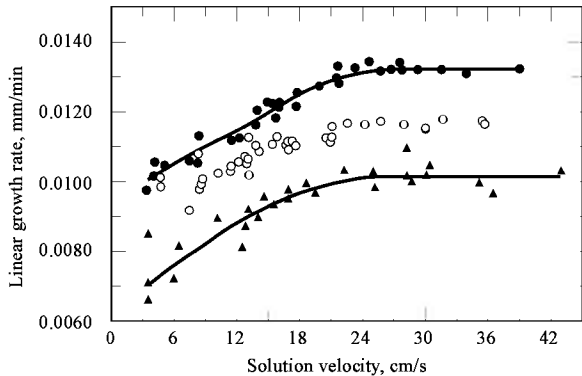


Fig. 8. Effect of solution velocity on the growth rate of the < 110 > face of MgSO₄·7H₂O. Concentration of MgSO₄: ●, 29.02% wt % at 30.8°C; ○, 28.89 wt % at 30.8°C; ▲, 18.89 wt % at 31.3°C.

Growth Models. Numerous models have been proposed to describe surface reaction kinetics, including those that assume crystals grow by layers and others that consider growth to occur by the movement of a continuous step. Each model results in a specific relationship between growth rate and supersaturation, but none can be used for a priori predictions of growth kinetics. Insights regarding the roles of certain process variables can be obtained, however, and with additional research predictive capabilities may be achieved. For these reasons and because of the extensive literature on the subject (5,24–26) all that will be pointed out here are the key aspects of the physical models and the resulting relationship between growth and supersaturation predicted by each theory.

Models used to describe the growth of crystals by layers call for a two-step process: (1) formation of a two-dimensional nucleus on the surface and (2) spreading of the solute from the two-dimensional nucleus across the surface. The relative rates at which these two steps occur give rise to the mononuclear two-dimensional nucleation theory and the polynuclear two-dimensional nucleation theory. In the mononuclear two-dimensional nucleation theory, the surface nucleation step occurs at a finite rate, whereas the spreading across the surface is assumed to occur at an infinite rate. The reverse is true for the polynuclear two-dimensional nucleation theory. From the mononuclear two-dimensional nucleation theory, growth is related to supersaturation by the equation.

$$G = C_1 h A [\ln(1 + s)]^{1/2} \exp \left[- \frac{C_2}{T^2 \ln(1 + s)} \right] \tag{16}$$

where C_1 and C_2 are system-dependent constants, h is the height of the nucleus, A is surface area, and s and T are as defined earlier. The polynuclear two-dimensional theory produces the equation

$$G = \left(\frac{C_3}{T^2 [\ln(1 + s)]^{3/2}} \right) \exp \left[- \frac{C_2}{T^2 \ln(1 + s)} \right] \tag{17}$$

where C_3 is a system-dependent constant. Finally, if both formation of the two-dimensional nucleus and spreading of the surface layer are important in determining growth rate, the following equation can be derived:

$$G = C_4 s^{2/3} [\ln(1 + s)]^{1/6} \exp \left[- \frac{C_2}{3T^2 \ln(1 + s)} \right] \tag{18}$$

where C_4 is a system-dependent constant.

Equation 16–18 can be simplified considerably by recognizing that in many systems the quantity s is much less than 1. In that case, $\ln(1 + s)$ is approximately s . Making this substitution, the growth rate from the mononuclear two-dimensional nucleation theory becomes

$$G = C_1 h A s^{1/2} \exp \left(- \frac{C_2}{T^2 s} \right) \tag{19}$$

For the polynuclear two-dimensional nucleation theory

$$G = \left(\frac{C_3}{T^2 s^{3/2}} \right) \exp \left(- \frac{C_2}{T^2 s} \right) \tag{20}$$

For both steps occurring at similar rates

$$G = C_4 s^{5/6} \exp \left(- \frac{C_2}{T^2 s} \right) \tag{21}$$

The screw dislocation theory (27), often referred to as the BCF theory (after its formulators), shows that the dependence of growth rate on supersaturation can vary from a parabolic relationship at low supersaturation to a linear relationship at high supersaturation. In the BCF theory, growth rate is given by

$$G = C \left(\frac{\epsilon s^2}{\sigma'_1} \right) \tanh \left(\frac{\sigma'_1}{\epsilon s} \right) \tag{22}$$

where ϵ is screw dislocation activity and σ'_1 is a system-dependent quantity that is inversely proportional to temperature. The dependence of growth rate on supersaturation is linear if the ratio $\sigma'_1 / \epsilon s$ is large, but the dependence becomes parabolic as the ratio becomes small. This is because $(\frac{\sigma'_1}{\epsilon s}) \rightarrow \frac{\sigma'_1}{\epsilon s}$ as $\frac{\sigma'_1}{\epsilon s}$ becomes small (supersaturation becomes large), and $(\frac{\sigma'_1}{\epsilon s}) \rightarrow 1.0$ as $\frac{\sigma'_1}{\epsilon s}$ becomes large (supersaturation becomes small). It is thus possible to observe variations in the dependence of growth rate on supersaturation for a given crystal-solvent system.

An empirical approach can also be used to relate growth kinetics to supersaturation with a power-law function of the form

$$G = k_G s^g \tag{23}$$

where k_G and g are constants determined by fitting the equation to growth-rate data. Such an approach should be valid over small ranges of supersaturation, and analysis of the theories discussed above shows that the more fundamental equations can be fit by equation 23 over limited ranges of supersaturation. For example, using the empirical approach to describe systems in which the screw dislocation model was applicable would limit g to values between 1 and 2, assuming ϵ was independent of supersaturation.

All the models described above indicate the importance of system temperature on growth rate. Dependencies of growth kinetics on temperature are often expressed in terms of an Arrhenius expression:

$$k_G = k_G^0 \exp \left(- \frac{\Delta E_G}{RT} \right) \tag{24}$$

where k_G is a growth rate coefficient of the type required in equation 23, k_G^0 is a constant, and ΔE_G is an activation energy. The magnitude of ΔE_G can be as large as that for many chemical reactions, 42 kJ/mol (>10 kcal/mol).

Both supersaturation and temperature can have different effects on the growth rates of different faces of the same crystal. Such occurrences have implications with respect to crystal habit, and these are dealt with in a later section.

Effects of Impurities and Solvent. The presence of impurities usually decreases the growth rates of crystalline materials, and problems associated with the production of crystals smaller than desired are commonly attributed to contamination of feed solutions. Strict protocols should be followed in operating units upstream from a crystallizer to minimize the possibility of such occurrences. Equally important is monitoring the composition of recycle streams so as to detect possible accumulation of impurities. Furthermore, crystallization kinetics used in scaleup should be obtained from experiments on solutions as similar as possible to those expected in the full-scale process.

The effects of a solvent on growth rates have been attributed to two sets of factors (28): one has to do with the effects of solvent on mass transfer of the solute through adjustments in viscosity, density, and diffusivity; the second is concerned with the structure of the interface between crystal and solvent. The analysis (28) concludes that a solute-solvent system that has a high solubility is likely to produce a rough interface and, concomitantly, large crystal growth rates.

Crystal Growth in Mixed Crystallizers. Multicrystal magma studies usually involve examination of the rate of change of a characteristic crystal dimension or the rate of increase in the mass of crystals. The characteristic dimension depends on the method used in the determination of size; eg, the second-largest dimension is measured by sieve analyses, whereas both electronic-zone-sensing and laser-light-scattering instruments provide estimates of an equivalent spherical diameter. If the rate of change of a crystal mass dM_c/dt is measured, the quantity can be related to the rate of change in the crystal characteristic dimension by the equation

$$\frac{dM_c}{dt} = \frac{d(\rho_c k_v L^3)}{dt} = 3\rho_c k_v L^2 \frac{dL}{dt} \tag{25}$$

where ρ_c is crystal density and k_v is the volume shape factor. Because an area shape factor can be defined by the equation

$$k_a = A_c L^2 \tag{26}$$

and G is defined as dL/dt ,

$$\frac{dM_c}{dt} = 3\rho_c \left(\frac{k_v}{k_a}\right) A_c G \tag{27}$$

The formulation of a population balance requires defining growth rate as the rate of change of the characteristic dimension

$$G = \frac{dL}{dt} \tag{28}$$

and solution of the resulting differential population balance requires a knowledge of the relationship between growth rate and size of the growing crystals. Moreover, this relationship can often be deduced from the form of population density data. A special condition, which simplifies such balances, results when all crystals in the magma grow at the same rate. Crystal-solvent systems that show this behavior are said to follow the ΔL law (29) whereas systems that do not are said to exhibit anomalous growth.

Anomalous growth means that growth rates of crystals in a magma are not identical or that the growth rate of an individual crystal or mass of crystals is not constant. Two theories have been used to explain growth rate anomalies: size-dependent growth and growth rate dispersion. Both alter the form of the population density function obtained from perfectly mixed continuous crystallizers; unfortunately, such behavior cannot be used to distinguish between size-dependent growth and growth rate dispersion, as both have the same qualitative effects on population density.

Size-dependent Crystal Growth. A number of empirical expressions correlate the apparent effect of crystal size on growth rate (30). The most commonly used correlation uses three empirical parameters to correlate growth rate with crystal size:

$$G = G^0 (1 + \gamma L)^b; \quad b < 1 \tag{29}$$

where G^0 , γ , and b are determined from experimental data. There have been attempts to relate the kinetic parameter b to crystallizer variables; the only success in this regard (31) showed a qualitative dependence on crystallizer volume. Several theories have been proposed to explain size-dependent growth kinetics, but none has been substantiated by direct observation or used to predict the onset of such behavior. One explanation seems particularly appealing: larger crystals impact impellers and other crystallizer internals with higher frequency and energy than smaller crystals; therefore, the larger crystals are recipients of more surface breaks and irregularities that lead to higher growth rates.

Growth Rate Dispersion. This phenomenon is the exhibition of different growth rates by crystals in a magma, even though they may have the same size and are exposed to identical conditions. It is now generally accepted that many observations originally attributed to size-dependent growth were due to growth rate dispersion. Such erroneous interpretations were the result of similarities in the effects of the two types of behavior on crystal size distributions.

The effects of growth rate dispersion on a population of sucrose crystals were first characterized by a linear correlation of the variance of the population about a mean size L with the extent of growth (30). It was demonstrated later (32,33) that growth rate dispersion could account for anomalous characteristics in the population density of crystals obtained from continuous, steady-state crystallizers. Later studies examined batch crystallization data to show that apparent size-dependent growth of potassium alum crystals was a manifestation of growth rate dispersion (34). Crystals of citric acid monohydrate generated by contact nucleation were found to exhibit growth rate dispersion but not size-dependent growth (35).

Two distinctly different mechanisms leading to growth rate dispersion have experimental support. The first assumes that all crystals have the same time-averaged growth rate, but the growth rates of individual crystals fluctuate about some mean value (36). Direct evidence of random fluctuations in growth rates has been reported for magnesium sulfate heptahydrate (37) and potassium alum (38). The second assumes that crystals are formed with a characteristic distribution of growth rates, but individual crystals retain a constant growth rate throughout their residence in a crystallizer. Findings on citric acid (35), potassium nitrate (39), and ammonium dihydrogen phosphate (40) support this mechanism.

Surface integration is thought to be the primary factor in both mechanisms of growth rate dispersion. The BCF theory indicates that the growth rate of a crystal face depends on the number, sign, and location of screw dislocations on the surface of a growing crystal. Collisions of crystals with each other and crystallizer internals result in changes in the dislocation network of a crystal, thereby leading to random fluctuations of growth rates. Changes in the dislocation networks also occur simply due to the imperfect growth of crystal faces. A distribution of growth rates is a result of the varying dislocation networks and densities among nuclei and seed crystals.

Although evidence exists for both mechanisms of growth rate dispersion, separate mathematical models were developed for incorporating the two mechanisms into descriptions of crystal populations: random growth rate fluctuations (36) and growth rate distributions (33,40). Both mechanisms can be included in a population balance to show the relative effects of the two mechanisms on crystal size distributions from batch and continuous crystallizers (41).

Crystal Characteristics

The morphology (including crystal shape or habit), size distribution, and purity of crystalline materials can determine the success in fulfilling the function of a crystallization operation.

Morphology. A crystal is highly organized, and constituent units, which can be atoms, molecules, or ions, are positioned in a three-dimensional periodic pattern called a space lattice. A characteristic crystal shape results from the regular internal structure of the solid with crystal surfaces forming parallel to planes formed by the constituent units. The surfaces (faces) of a crystal may exhibit varying degrees of development, with a concomitant variation in macroscopic appearance.

If atoms, molecules, or ions of a unit cell are treated as points, the lattice structure of the entire crystal can be shown to be a multiplication in three dimensions of the unit cell. Only 14 possible lattices (called Bravais lattices) can be drawn in three dimensions. These can be classified into seven groups based on their elements of symmetry. Moreover, examination of the elements of symmetry (about a point, a line, or a plane) for a crystal shows that there are 32 different combinations (classes) that can be grouped into seven systems. The correspondence of these seven systems to the seven lattice groups is shown in Table 1.

Table 1. The 14 Bravais Lattices^a

Type of symmetry	Lattice	Crystal system
cubic	cube	regular
	body-centered cube	
	face-centered cube	
tetragonal	square prism	tetragonal
	body-centered square prism	
orthorhombic	rectangular prism	orthorhombic
	body-centered rectangular prism	
	rhombic prism	
	body-centered rhombic prism	
monoclinic	monoclinic parallelepiped	monoclinic
	clinorhombic	
	triclinic parallelepiped	
triclinic	triclinic parallelepiped	triclinic
rhomboidal	rhombohedron	triclinic
hexagonal	hexagonal prism	hexagonal

^a Ref. 5.

The general shape of a crystal is referred to as its habit. The appearance of the crystalline product and its processing characteristics (such as washing and filtration) are affected by crystal habit. Relative growth rates of the faces of a crystal determine its shape; faster-growing faces become smaller than slower-growing faces and, in the extreme, may disappear from the crystal altogether. Growth rates depend on the presence of impurities, rates of cooling, temperature, solvent, mixing, and supersaturation. Furthermore, the importance of each of these factors may vary from one crystal face to another. For example, consider Figure 9 which shows that the $\langle 111 \rangle$ face grows between 1.6 and 2.2 times as fast as the $\langle 110 \rangle$ face at the conditions examined. These results account for the elongated crystal shape exhibited by magnesium sulfate heptahydrate crystals. In addition, the effects of supersaturation and temperature are different on the growth rates of the two faces studied. Such behavior leads to changes in habit as the temperature and or supersaturation are changed in a crystallizer.

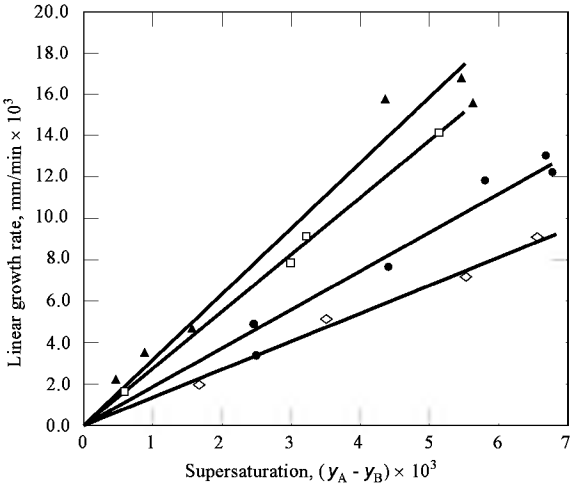
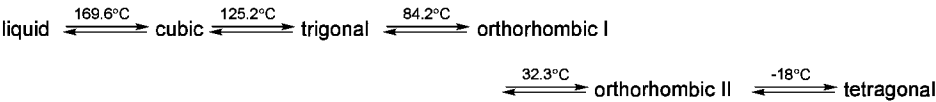


Fig. 9. Growth rate of faces of $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$. $\blacktriangle$, $\langle 111 \rangle$ face at $T_{\text{sat}} = 38.0^\circ\text{C}$; $\blacksquare$, $\langle 111 \rangle$ face at $T_{\text{sat}} = 35.5^\circ\text{C}$; $\bullet$, $\langle 110 \rangle$ face at $T_{\text{sat}} = 38.0^\circ\text{C}$; $\blacklozenge$, $\langle 110 \rangle$ face at $T_{\text{sat}} = 33.5^\circ\text{C}$; solution velocity = 1.2 cm/s (42). Reprinted with permission of the American Chemical Society.

A number of studies have shown that various additives can be included in a process stream to alter crystal habit (5). Prediction of such behavior is difficult and extensive laboratory or bench-scale experiments may be required to evaluate the effectiveness of habit modifiers. More recently, some measure of success has been achieved with altering the habit of organic crystals based on the molecular structure and characteristics of the crystallizing species. One category of additives affecting crystal habit is surfactants (43). Should an additive enhance the properties of a crystalline material, for example, by making it easier to filter, the expense associated with its use may be warranted. Significant efforts toward tailoring additives so that they have specific effects on crystal habit have been made by a number of research groups (44,45).

Polymorphism is a condition in which a specific chemical substance may crystallize into different forms. For example, ammonium nitrate exhibits four changes in form (5) between -18° and 125°C :



Transitions from one polymorphic form to another may be accompanied by changes in specific volume, which may lead to destruction of the crystal and containers in which the substance is stored.

A specific polymorph may be absolutely essential for a crystalline product, for example, one polymorph may have a more desirable color or greater hardness or disperse in water more easily than another polymorph. Often, one polymorphic form is more stable than another (for example, at 80°C the orthorhombic I form of ammonium nitrate is more stable than the trigonal form) at conditions to which a product is exposed. An interesting approach to

keeping a less stable polymorph from transforming to a more stable, but less desirable, polymorphic form uses an additive to block rearrangement of the molecular structure leading to the undesired form (46).

Agglomeration. Many of the analyses of industrial crystallizers require that the particle recovered from the crystallizer consist of a single crystal. It is only with this type of system that single growth rates are likely to be exhibited and, moreover, many of the properties of the crystal are affected deleteriously by agglomeration. Purity, for example, typically is diminished when agglomeration occurs. Countering the negative aspects of agglomeration is recognition that in many systems the single crystals produced by normal crystal growth would be too small to be separable using conventional solid–liquid separation equipment. In such instances, there would be no product without agglomeration.

The needs associated with a greater understanding of agglomeration fall into at least two separate areas: identifying the variables affecting agglomeration and accounting properly for agglomeration in a population balance around a crystallizer. The latter of these has been addressed (47,48), but the subject matter is considered beyond the scope of the present discussion.

Several process variables that influence agglomeration of copper sulfate pentahydrate [7758-99-8] crystals have been identified (49). Particles originally appearing to be single crystals were found to be agglomerates of complex structure. Electron photomicrographs showed that the agglomerates were formed early in the life of crystals comprising the agglomerates and that the growth of these agglomerates had followed a complex and unpredictable pattern. The percentages of agglomerates in the particles recovered from a series of mixed-suspension, mixed-product experiments were found to increase with increasing supersaturation, increasing magma density, and decreasing agitation. The observations fit with a hypothesis that the agglomeration resulted when two or more crystals came together and were bonded through overgrowth of contact areas. Such a hypothesis is inadequate, however, in predicting when agglomeration will occur and the key variables that can be adjusted to control the agglomeration.

Purity. Although crystallization has been employed extensively as a separation process, purification techniques using crystallization have become increasingly important. Mechanisms by which impurities can be incorporated into crystalline products include adsorption of impurities on crystal surfaces (50), solvent entrapment in cracks, crevices and agglomerates, and inclusion of pockets of liquid (51). An impurity having a structure sufficiently similar to the material being crystallized can also be incorporated into the crystal lattice by substitution or entrapment (52,53). Among these mechanisms, inclusion formation has been extensively studied (54–58). It has also been suggested that the purity may be directly linked to size and habit of product crystals, but the interaction does not appear to be simple. It has been noted (59) that the key to producing high purity crystals was to maintain the supersaturation at a low level so that large crystals were obtained. Others have found that reducing the size of ammonium perchlorate crystals resulted in a substantial decrease in moisture due to inclusion (58).

Key factors in solving problems associated with crystal purity are the location, ie, on the surface or incorporated in the crystal, and the nature of the impurity. Impurities are on the exterior of host crystals as a result of adsorption; wetting by a solvent that contains the impurities; or entrapment of impure solvent in cracks, crevices, agglomerates, and aggregates. Incorporation of impurities within crystals comes about through formation of inclusions (also referred to as occlusions) of solvent, lattice substitution, or lattice entrapment. Obviously, the characteristics of an impurity determine whether it is positioned on the surface or the interior of host crystals. Three key impurity types are solutes similar to the product, solutes dissimilar from the product, and the solvent. The effects of process variables on the purity of L-isoleucine crystallized from aqueous solutions containing other amino acids (impurities) have been determined (60). Another study has examined how the methanol content of L-serine crystals could be minimized when the mode of crystallization is addition of methanol to saturated aqueous solutions (4). Mixing and the rate at which supersaturation is generated are important in both of these cases.

Crystal Size Distributions. Particulate matter produced by crystallization has a distribution of sizes that varies in a definite way over a specific size range. A crystal size distribution (CSD) is most commonly expressed as a population (number) distribution relating the number of crystals at each size to the size or as a mass (weight) distribution expressing how mass is distributed over the size range. The two distributions are related and affect many aspects of crystal processing and properties, including appearance, solid–separation, purity, reactions, dissolution, and other properties involving surface area.

Population density (n) has dimensions number (volume)(length); it is a key quantity in the discussion of CSD, a function of the characteristic crystal dimension L , and it is defined so that it is independent of the magnitude of the system. When a total population density is used, the symbol is $\bar{n}$ and the units are number length. Population density is defined by letting ΔN be the number of crystals per unit system volume in a size range from L to $L + \Delta L$, so that

$$n = \lim_{\Delta L \rightarrow 0} \frac{\Delta N}{\Delta L} \tag{30}$$

The arbitrary system volume on which n is based must be defined before the population density function has meaning. For example, the volume may be that of the slurry or the clear liquor in the system.

The function N in equation 30 is a cumulative number distribution representing the number of crystals per unit volume in the distribution that have a characteristic dimension less than L' . Therefore,

$$N(L') = \int_0^{L'} n dL \tag{31}$$

and the fraction of the crystals in the distribution $F(L')$ that have a size less than L' can be calculated as

$$F(L') = \frac{N(L')}{N_{\text{tot}}} \tag{32}$$

A mass (weight) density function, given the symbol m and having dimensions mass (volume)(length), can be defined analogously to population density; letting ΔM be the mass of crystals per unit system volume in the size range L to $L + \Delta L$,

$$m = \lim_{\Delta L \rightarrow 0} \frac{\Delta M}{\Delta L} \tag{33}$$

The two density functions can be related through a simple shape factor as follows. Suppose the mass of a single crystal is M_c and the characteristic dimension of that crystal is L . If the crystal is from a population in which shape is not a function of size, then the mass of any crystal from that population is related to characteristic dimension by a volume shape factor:

$$M_c = k_v \rho L^3 \tag{34}$$

Recognizing that the mass of crystals in a sample is the product of the number of crystals and the mass of a single crystal, mass and population densities may be related by the expression

$$\Delta M = k_v \rho L^3 \Delta N \tag{35}$$

where ρ is crystal density. Dividing by ΔL and taking the limit as this quantity approaches zero,

$$m = k_v \rho L^3 n$$

(36)

The function M used in the above equations is a cumulative mass distribution function, representing the mass of crystals having a characteristic dimension less than L' . The total mass of crystals per unit volume is related to population density by the equation

$$M_T = k_v \rho \int_0^\infty L^3 n dL$$

(37)

This quantity, which is often referred to as magma density or solids concentration (mass of crystals per unit system volume), is often an important process variable. A cumulative mass fraction of crystals having a size less than L' can also be defined as

$$W(L') = \frac{M(L')}{M_T} = \frac{k_v \rho \int_0^{L'} L^3 n(L) dL}{M_T}$$

(38)

Moments of a distribution often provide information that can be used to characterize particulate matter. The j th moment of the population density function n is defined as

$$m_j = j\text{th moment} = \int_0^\infty L^j n dL$$

(39)

It can be demonstrated that the total number of crystals, the total length, the total area, and the total volume of crystals, all in a unit of system volume, can be evaluated from the zero, first, second, and third moments of the population density function.

An average crystal size can be used to characterize a CSD. However, the average can be determined on any of several bases, and the basis selected must be specified for the average to be useful. More than 20 different averaging procedures have been proposed, yet none is generally satisfactory or preferred (5).

The complete characterization of a particulate material requires development of a functional relationship between crystal size and population or mass. The functional relationship may assume an analytical form (7), but more frequently it is necessary to work with data that do not fit such expressions. As such detail may be cumbersome or unavailable for a crystalline product, the material may be more simply (and less completely) described in terms of a single crystal size and a spread of the distribution about that specified dimension.

The dominant crystal size, L_D , is most often used as a representation of the product size, because it represents the size about which most of the mass in the distribution is clustered. If the mass density function defined in equation 33 is plotted for a set of hypothetical data as shown in Figure 10, it would typically be observed to have a maximum at the dominant crystal size. In other words, the dominant crystal size L_D is that characteristic crystal dimension at which $dm/dL = 0$. Also shown in Figure 10 is the theoretical result obtained when the mass density is determined for a perfectly mixed, continuous crystallizer within which invariant crystal growth occurs. That is, mass density is found for such systems to follow a relationship of the form $m = aL^3 \exp(-bL)$, where a and b are system-dependent parameters.

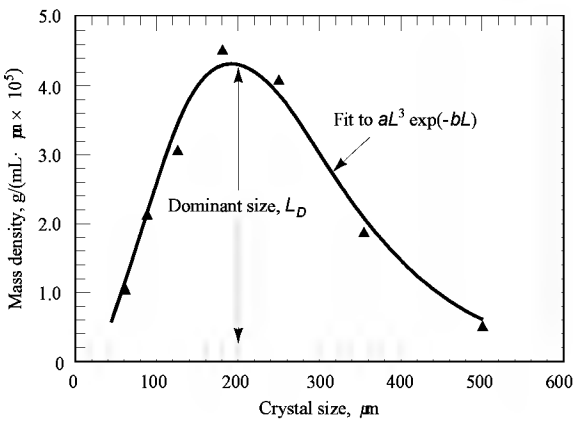


Fig. 10. Determination L_D from plot of mass density function.

The coefficient of variation (cv) of a distribution is a measure of the spread of the distribution about some characteristic size. It is often used in conjunction with dominant size to characterize crystal populations through the equation

$$cv = \frac{\sigma}{L_D}$$

(40)

where σ is the standard deviation of the distribution. The coefficient of variation of the mass density function about the dominant crystal size is given by

$$cv = \left(\frac{m_3 m_5}{m_4^2} - 1 \right)^{1/2}$$

(41)

Population Balances and Crystal Size Distributions

Population balances and crystallization kinetics may be used to relate process variables to the crystal size distribution produced by the crystallizer. Such balances are coupled to the more familiar balances on mass and energy. It is assumed that the population distribution is a continuous function and that crystal size, surface area, and volume can be described by a characteristic dimension L . Area and volume shape factors are assumed to be constant, which is to say that the morphology of the crystal does not change with size.

A balance is formulated around a control volume V_T on the number of crystals in any size range, say L_1 to L_2 . It must account for crystals that enter and leave the size range by convective flow and crystals that enter and leave the size range by crystal growth. Crystal breakage and agglomeration are assumed to be negligible in the present analysis, and it is assumed that crystals are formed by nucleation at size zero. The rate of crystal growth G is defined as the rate of change of the characteristic crystal dimension L ; that is, $G = dL/dt$.

Consider the crystallizer shown in Figure 11. If it is assumed that the crystallizer is well mixed with a constant slurry volume V_T then, as shown (7), the following partial differential population balance can be derived:

$$\frac{\partial(nG)}{\partial L} + \frac{Q_0 n}{V_T} - \frac{Q_i n_i}{V_T} = \frac{\partial n}{\partial t}$$

(42)

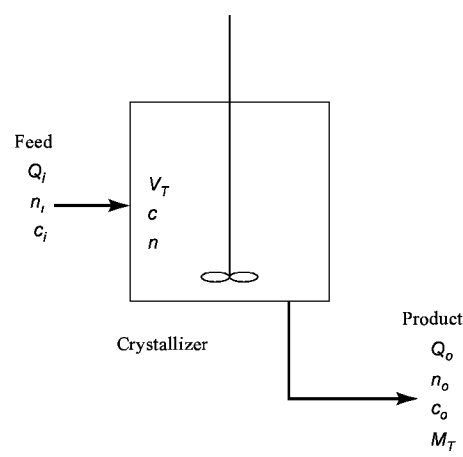


Fig. 11. Schematic diagram of a simple, perfectly mixed crystallizer.

If the crystallizer is now assumed to operate with a clear feed ($n_i = 0$), at steady state ($\partial n / \partial t = 0$), and if the crystal growth rate G is invariant and a mean residence time τ is defined as V_T/Q_0 , then the population balance can be written as

$$G \frac{dn}{dL} + \frac{n}{\tau} = 0$$

(43)

τ is often referred to as the drawdown time to reflect that it is the time required to empty the contents from the crystallizer if the feed is set to zero. Equation 43 can be integrated using the boundary condition $n = n^0$ at $L = 0$:

$$n = n^0 \exp\left(-\frac{L}{G\tau}\right)$$

(44)

If the magma volume V_T is allowed to vary in the system on which equation 42 is based, the population balance becomes

$$\frac{\partial n}{\partial t} + \frac{\partial(nG)}{\partial L} + n \frac{\partial(\ln V_T)}{\partial t} + \frac{Q_0 n}{V_T} = 0$$

(45)

The crystallizer model that led to the development of equations 44 and 45 is referred to as the mixed-suspension, mixed-product removal (MSMPR) crystallizer.

Determination of Crystallization Kinetics. Under steady-state conditions, the total number production rate of crystals in a perfectly mixed crystallizer is identical to the nucleation rate, B^0 . Accordingly,

$$B^0 = \frac{1}{\tau} \int_0^\infty n dL$$

(46)

For crystallizers following the constraints leading to equation 44,

$$B^0 = n^0 G$$

(47)

Combining equations 45 and 48

$$n = \frac{B^0}{G} \exp\left(-\frac{L}{G\tau}\right)$$

(48)

Analysis of equation 48 shows that a single sample taken either from inside the crystallizer or from the product stream will allow evaluation of nucleation and growth rates at the system conditions. Figure 12 shows a plot of typical population density data obtained from a crystallizer meeting the stated assumptions. The slope of the plot of such data may be used to obtain the growth rate, and the product of the intercept and growth rate gives the nucleation rate.

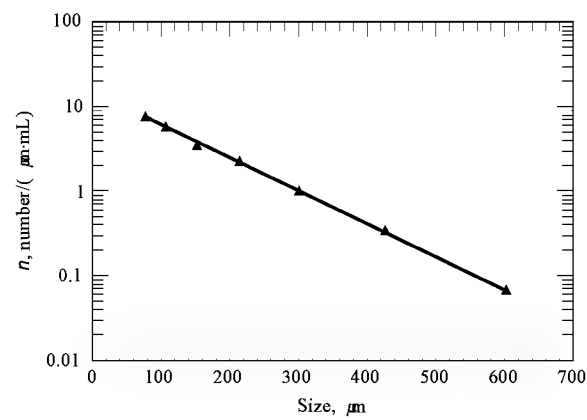


Fig. 12. Plot of population density as a function of size for KNO_3 , $\tau = 15$ min. For the line, $n = 16.528 \cdot \exp(-0.0090426L)$; $R = 0.99752$; slope $= -1/G\tau$;

intercept = $n^0 = B^0 / G$.

Many industrial crystallizers operate in a well-mixed or nearly well-mixed manner, and the equations derived above can be used to describe their performance. Furthermore, the simplicity of the equations describing an MSMPR crystallizer make experimental equipment configured to meet the assumptions leading to equation 44 useful in determining nucleation and growth kinetics in systems of interest.

From a series of runs at different operating conditions, a correlation of nucleation and growth kinetics with appropriate process variables can be obtained; the resulting correlation can then be used to guide either crystallizer scaleup or the development of an operating strategy for an existing crystallizer. The variables affecting nucleation and growth kinetics include temperature, supersaturation, magma density, and external stimuli, such as agitation or circulation rate of the magma. Empirical power-law functions are used most frequently in correlating nucleation and growth rates, but the choice of the independent variables can be justified from a mechanistic perspective. For example, systems that are believed to follow secondary nucleation mechanisms should include a variable such as magma density that reflects the character of crystals in the crystallizer. The most commonly used power-law functions are

$$B^0 = k_1 s^b M_T^j \tag{49}$$

$$G = k_2 s^g \tag{50}$$

It is often difficult to measure supersaturation, especially in systems that have high growth rates. Even though the supersaturation in such systems is so small that it can be neglected in writing a solute mass balance, it is important in setting nucleation and growth rates. In such instances it is convenient to substitute growth rate for supersaturation by combining equations 49 and 50 to get

$$B^0 = k_n G^i M_T^j \tag{51}$$

The constant k_n may depend on process variables such as temperature, rate of agitation or circulation, presence of impurities, and other variables. If sufficient data are available, such quantities may be separated from the constant by adding more terms in a power-law correlation. The term k_n is specific to the operating equipment and generally is not transferrable from one equipment scale to another. The system-specific constants i and j are obtainable from experimental data and may be used in scaleup, although j may vary considerably with mixing conditions. Illustration of the use of data from a commercial crystallizer to obtain the kinetic parameters k_n , i , and j is available (61).

Mass Balance Constraints. From the schematic diagram of a continuous crystallizer shown in Figure 11, the following mass balance on solute can be constructed:

$$Q_i c_i = Q_0 c_0 + Q_T M_T \tag{52}$$

c_0 is determined by system kinetics and constrained by a solid–liquid equilibrium (solubility) relationship, which gives the equilibrium concentration c^* at the system conditions. The system (solute–solvent and crystallizer) is characterized by the magnitude of the supersaturation ($c_0 - c^*$) remaining in the solution exiting the crystallizer. If the mass balance is closed by substituting c^* for c_0 in equation 52, the system is said to be a fast-growth or class II system. If the mass balance is not closed, significant supersaturation remains in the solution and the system is said to be a slow-growth or class I system. In other words, for class I (slow-growth) systems: $c_0 > c^*$ and

$$M_T = \frac{Q_i}{Q_0} c_i - c_0 \tag{53}$$

Values of process variables such as residence time may change the system kinetics that change c_0 and, in turn, M_T .

For class II (fast-growth) systems: $c_0 = c^*$ and

$$M_T = \frac{Q_i}{Q_0} c_i - c^* \tag{54}$$

Process variables do not change c^* and, therefore, M_T is constant over modest ranges of operating conditions.

CSD Characteristics for MSMPR Crystallizers. The perfectly mixed crystallizer described in the preceding discussion is highly constrained and the form of crystal size distributions produced by such systems is fixed. Such distributions have the following characteristics.

(1) Moments of the distribution can be calculated for MSMPR crystallizers by the simple expression

$$m_j = j! n^0 (G\tau)^{j+1} \tag{55}$$

Properties of the distribution such as total number of crystals per unit volume, total length of crystals per unit volume, total area of crystals per unit volume, and total volume of solids (crystals) per unit volume may be explicitly evaluated from the moment equations.

(2) The dominant crystal size L_D is given by $L_D = 3G\tau$. This quantity is also the ratio m_3 / m_2 , which is often given the symbol $\bar{L}_{3,2}$.

(3) From the definition of the coefficient of variation given by equation 41, $cv = 50\%$ for an MSMPR crystallizer. Such a cv may be too large for certain commercial products, which means either the crystallizer must be altered or the product must be screened to separate the desired fraction.

(4) The magma density M_T (mass of crystals per unit volume of slurry or liquor) may be obtained from the third moment of the population density function and is given by

$$M_T = 6\rho k_v n^0 (G\tau)^4 \tag{56}$$

Although magma density is a function of the kinetic parameters n^0 and G , it often can be measured independently. In such cases, it should be used as a constraint in evaluating nucleation and growth rates from measured crystal size distributions (62), especially if the system of interest exhibits the characteristics of anomalous crystal growth.

(5) A pair of kinetic parameters, one for nucleation rate and another for growth rate, describe the crystal size distribution for a given set of crystallizer operating conditions. Variation in one of the kinetic parameters without changing the other is not possible. Accordingly, the relationship between these parameters determines the ability to alter the characteristic properties (such as dominant size) of the distribution obtained from an MSMPR crystallizer (7).

Preferential Removal of Crystals. Crystal size distributions produced in a perfectly mixed continuous crystallizer are highly constrained; the form of the CSD in such systems is determined entirely by the residence time distribution of a perfectly mixed crystallizer. Greater flexibility can be obtained through introduction of selective removal devices that alter the residence time distribution of materials flowing from the crystallizer. The

functions of classified removal are best described in terms of idealized models of clear-liquor advance, classified-fines removal, and classified-product removal.

Clear-liquor advance is simply the removal of mother liquor from the crystallizer without simultaneous removal of crystals. The primary objective of *fines removal* is preferential withdrawal from the crystallizer of crystals whose size is below some specified value. Such crystals may be redissolved and the resulting solution returned to the crystallizer. *Classified-product removal* is carried out to remove preferentially those crystals whose size is larger than some specified value.

The effects of each selective removal function on CSD can be described in terms of the population density function n . It is convenient to define flow rates in terms of clear liquor, which requires the population's density function to be defined on a clear-liquor basis. In the present discussion, only systems exhibiting invariant crystal growth are considered.

Clear-liquor advance reduces the quantity of liquor that must be processed by solid-liquid separation equipment (for example, a filter or a centrifuge). The reduction in liquor flow through the separation equipment may allow use of smaller equipment for a fixed production rate or increased production through fixed equipment.

The function of clear-liquor advance can be illustrated by considering a simple operation, shown in Figure 13, in which Q_i , Q_{CL} , and Q_o represent volumetric flow rates of clear-liquor fed to the crystallizer, in the clear-liquor advance, and in the output slurry. In such systems the population density function is given by the expression

$$n = n^0 \exp\left(-\frac{L}{G\tau_p}\right)$$

(57)

where $\tau_p = V/Q_o$. It is clear that increasing Q_{CL} decreases Q_o and thereby increases the residence time of the crystals in the crystallizer.

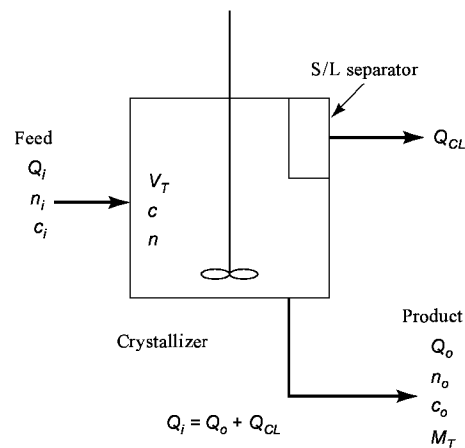


Fig. 13. Simplified schematic diagram of clear-liquor advance or double-draw off (DDO).

Clearly, the form of the population density function resulting from a clear-liquor advance system is identical to that expected from perfectly mixed systems in which τ_{crystals} is identical to τ_{liquor} . Unless the increase in magma density associated with clear-liquor advance results in significant increases in nucleation, some increase in the dominant crystal size can be expected. It has been observed that the increase in L_D may be greater than predicted from theory. This is caused by the stream being removed as clear liquor containing varying amounts of fines, which means the system characteristics are those of classified-fines removal.

As an idealization of the classified-fines removal operation, assume that two streams are withdrawn from the crystallizer, one corresponding to the product stream and the other a fines removal stream. Such an arrangement is shown schematically in Figure 14. The flow rate of the clear solution in the product stream is designated Q_o and the flow rate of the clear solution in the fines removal stream is set as $(R - 1)Q_o$. Furthermore, assume that the device used to separate fines from larger crystals functions so that only crystals below an arbitrary size L_F are in the fines removal stream and that all crystals below size L_F have an equal probability of being removed in the fines removal stream. Under these conditions, the crystal size distribution is characterized by two mean residence times, one for the fines and the other for crystals larger than L_F . These quantities are related by the equations

$$\tau_F = \frac{V}{RQ_o} = \frac{\tau}{R} \quad (L < L_F)$$

(58)

$$\tau = \frac{V}{Q_o} \quad (L > L_F)$$

(59)

where V is the volume of clear solution in the crystallizer. Recognize that the ratio of the probability of a crystal smaller than L_F being removed from the crystallizer to that of crystals larger than L_F being removed is $R = \tau / \tau_F$.

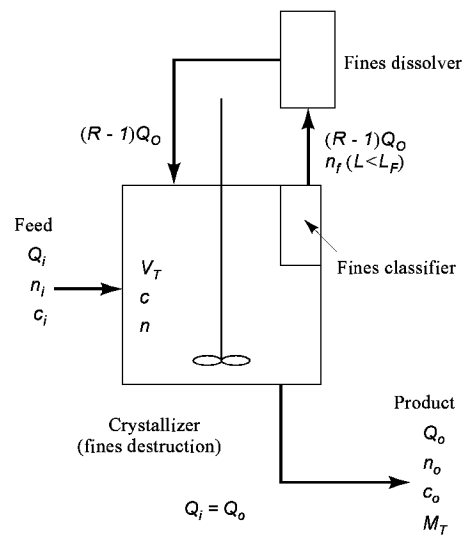


Fig. 14. Simplified schematic diagram of classified-fines removal.

For systems following invariant growth the crystal population density in each size range decays exponentially with the inverse of the product of growth rate and residence time. For a continuous distribution, the population densities of the classified fines and the product crystals must be the same at size L_F . Accordingly, the population density for a crystallizer operating with classified-fines removal is given by

$$n = n^0 \exp \left[\frac{RL}{G\tau} \right] \quad (L < L_F) \tag{60}$$

$$n = n^0 \exp \left[\frac{(R-1)L_F}{G\tau} \right] \exp \left[\frac{L}{G\tau} \right] \quad (L > L_F) \tag{61}$$

Figure 15 shows how the population density function changes with the addition of classified-fines removal. It is apparent from the figure that fines removal increases the dominant crystal size, but it also increases the spread of the distribution.

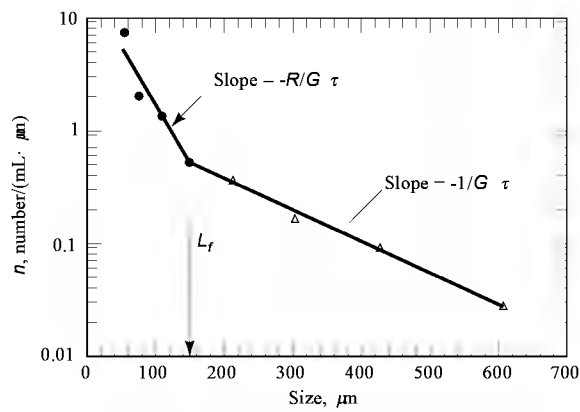


Fig. 15. Population density function for product from crystallizer with classified-fines removal. Cut size $L_F = 150 \mu\text{m}$; $R = 3.7$.

A simple method for implementation of classified-fines removal is to remove slurry from a settling zone in the crystallizer. The settling zone can be created by constructing a baffle that separates the zone from the well-mixed portion of the vessel, as is the case for a draft-tube-baffle crystallizer, or in small-scale systems, by simply inserting a length of pipe into the crystallizer chamber. The separation of crystals in the settling zone is based on the dependence of settling velocity on crystal size; only those crystals having a settling velocity greater than the upward velocity of the slurry remain in the crystallizer. As the cross-sectional area of a settling zone is invariant, the flow rate of slurry through the zone determines the cut size L_F and it also determines the parameter R used in equations 59–62.

Classified removal of course material also can be used, as shown in Figure 16. In a crystallizer equipped with idealized classified-product removal, crystals above some size L_C are removed at a rate Z times the removal rate expected for a perfectly mixed crystallizer, and crystals smaller than L_C are not removed at all. Larger crystals can be removed selectively through the use of an elutriation leg, hydrocyclones, or screens. Using the analysis of classified-fines removal systems as a guide, it can be shown that the crystal population density within the crystallizer magma is given by the equations

$$n = n^0 \exp \left[\frac{L}{G\tau} \right] \quad (L < L_C) \tag{62}$$

$$n = n^0 \exp \left[\frac{(Z-1)L_C}{G\tau} \right] \exp \left[\frac{ZL}{G\tau} \right] \quad (L > L_C) \tag{63}$$

where τ is defined as the residence time V/Q_o . Figure 17 shows the effects of classified-product removal on crystal size distribution. The characteristics of the crystal size distribution obtained from a system with classified product removal show a narrower distribution (reduced coefficient of variation) and smaller dominant size. A more complete discussion of the implications of classified-product removal, with particular attention given to the distinction between the crystal population densities within the crystallizer and in the product has been given (7).

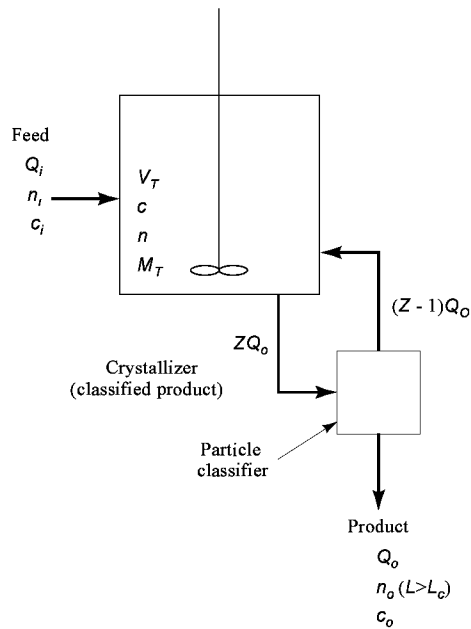


Fig. 16. Classified withdrawal of coarse crystals.

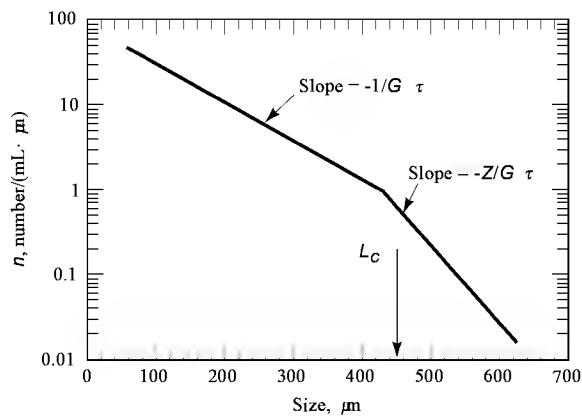


Fig. 17. Effect of classified-product removal on population densities within a crystallizer and in the crystallizer product.

It is possible to obtain both a narrowing of the distribution and an increase in dominant size by combining preferential removal of fines and coarse crystals. Including the idealized removal functions for fines and coarse crystals in a population balance and assuming invariant crystal growth will result in a population density function within the crystallizer given by equations 64–66, Figure 18 illustrates the effects of both removal functions on population density. This plot of population density results from sampling the magma within a crystallizer, not from sampling the product stream.

$$n = n^0 \exp \left[\frac{RL}{G\tau} \right] \quad \text{for } L \leq L_F \tag{64}$$

$$n = n^0 \exp \left[\frac{(R-1)L_F}{G\tau} \right] \exp \left[\frac{L}{G\tau} \right] \quad \text{for } L_F < L < L_C \tag{65}$$

$$n = n^0 \exp \left[\frac{(R-1)L_F}{G\tau} \right] \exp \left[\frac{(Z-1)L_C}{G\tau} \right] \exp \left[\frac{ZL}{G\tau} \right] \quad \text{for } L > L_C \tag{66}$$

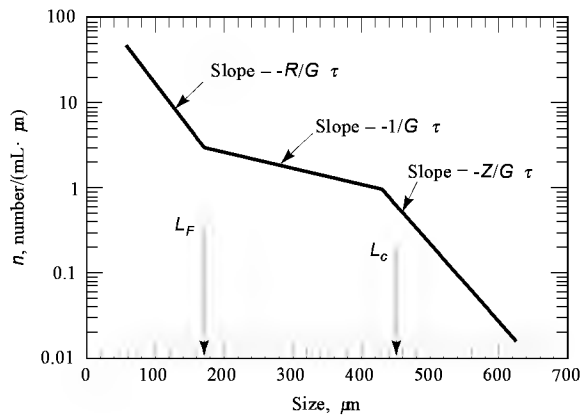


Fig. 18. Population density functions of crystals within a crystallizer, having both classified-fines and classified-product removal and of crystals in the product from such a crystallizer.

The model of the crystallizer and selective removal devices that led to equations 64–66 is referred to as the R-Z crystallizer. It is an obvious idealization of actual crystallizers because of the perfect cuts assumed at L_F and L_C . However, it is a useful approximation to many systems and it allows qualitative analyses of complex operations. The R-Z model may also be representative of inadvertant classification, ie, fines or course crystals may be preferentially removed from a crystallizer without installation of specific hardware to accomplish such an objective.

Although many commercial crystallizers operate with some form of selective crystal removal, such devices can be difficult to operate because of fouling of heat exchanger surfaces or blinding of screens. In addition, several investigations identify interactions between classified fines and course product removal as causes of cycling of a crystal size distribution (7). Often such behavior can be minimized or even eliminated by increasing the fines removal rate (63,64).

Batch Crystallization. Crystal size distributions obtained from batch crystallizers are affected by the mode used to generate supersaturation and the rate at which supersaturation is generated. For example, in a cooling mode there are several avenues that can be followed in reducing the temperature of the batch system, and the same can be said for the generation of supersaturation by evaporation or by addition of a nonsolvent or precipitant. The complexity of a batch operation can be illustrated by considering the summaries of seeded and unseeded operations shown in Figure 19.

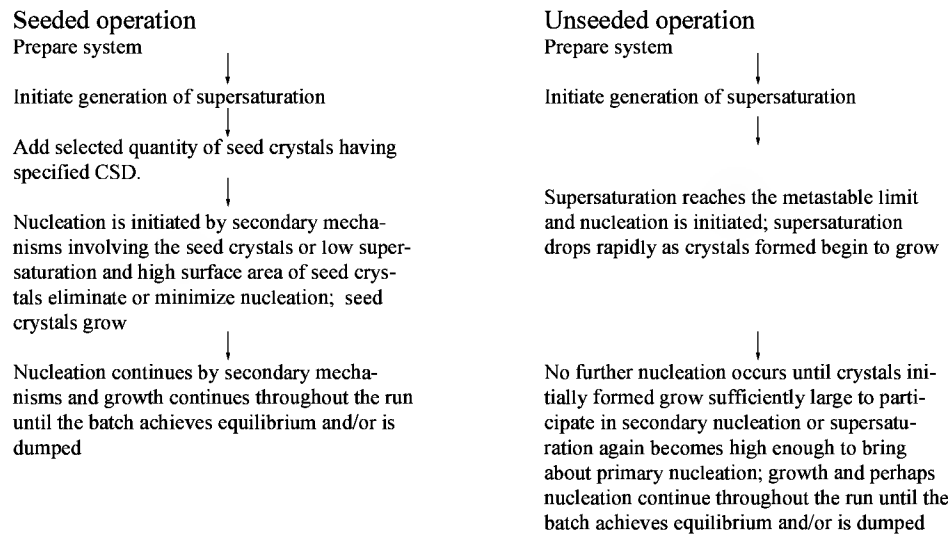


Fig. 19. Batch crystallizer operation.

The crystal size distributions resulting from the operating strategies outlined in Figure 19 depend greatly on the use of seeding, the rate at which supersaturation is generated, and those variables that are important in the prevailing mechanism of nucleation. Figures 20 and 21 summarize the qualitative variations in CSD that may be observed in batch crystallization and the role of adding seed crystals to such systems.

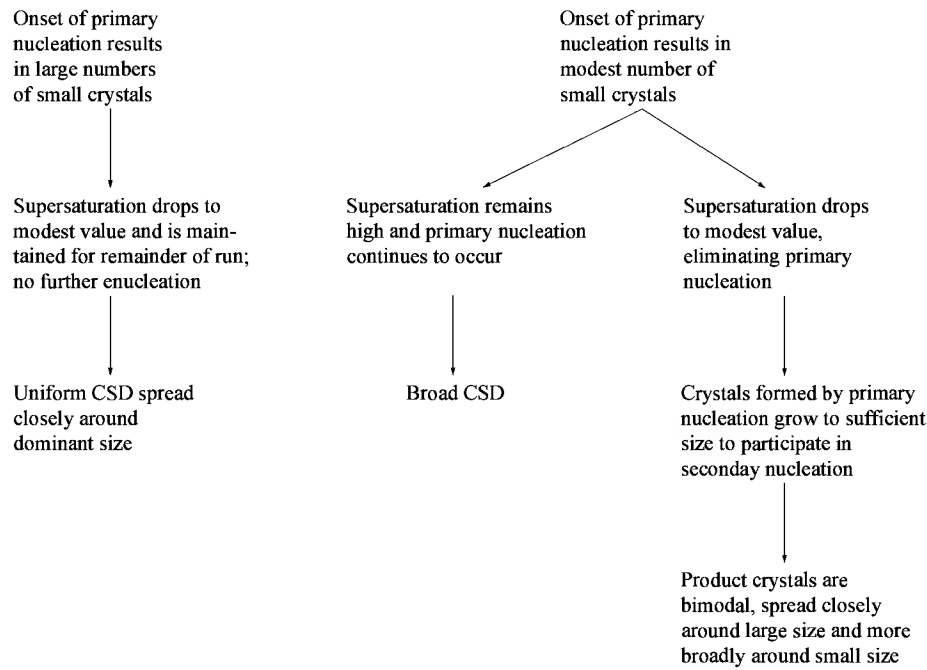


Fig. 20. CSD characteristics from batch crystallization without seeding.

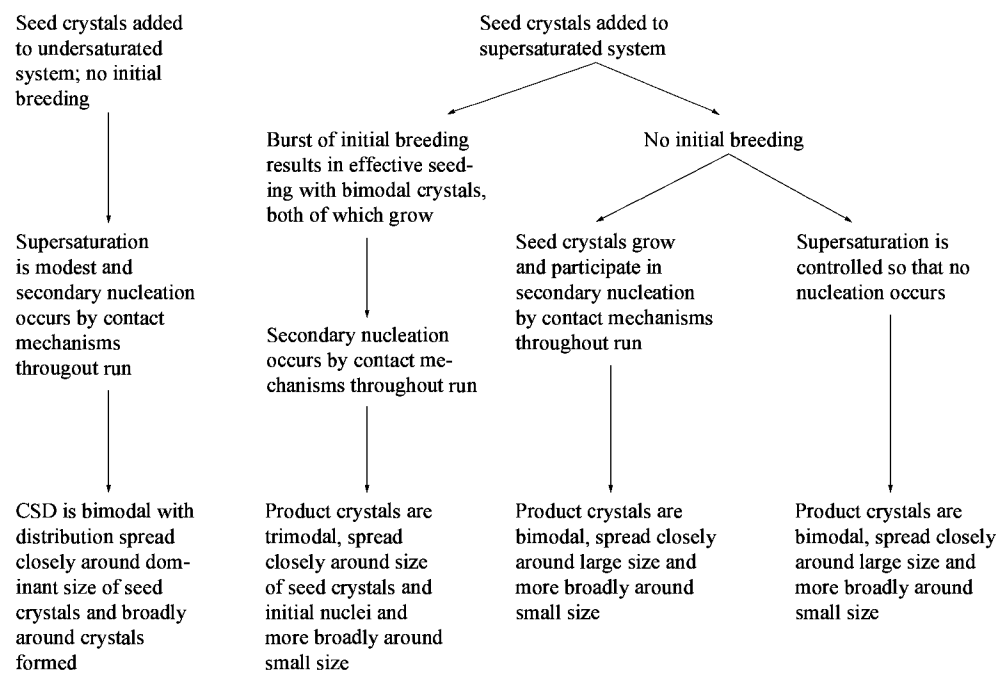


Fig. 21. CSD characteristics from batch crystallization with seeding.

More quantitative relationships of the CSD obtained from batch operations can be developed through formulation of a population balance. Using a population density defined in terms of the total crystallizer volume rather than on a specific basis ($\bar{n} = nV$), the general population balance given by equation 42 can be modified in recognition of there being no feed or product streams:

$$\frac{\partial(nV)}{\partial t} + \frac{\partial(GnV)}{\partial L} - \frac{\partial \bar{n}}{\partial t} + \frac{\partial(G\bar{n})}{\partial L} = 0 \tag{67}$$

The solution to this equation requires an initial condition ($\bar{n}$ at $t = 0$) and a boundary condition ($\bar{n}$ at a specific value of L). Assuming that crystals are formed at zero size gives the boundary condition:

$$n(0, t) = n^0(t) = \frac{B^0(t)}{G(0, t)} \tag{68}$$

Identification of an initial condition is difficult because of the problem of specifying the size distribution at the instant nucleation occurs. The difficulty is mitigated through the use of seeding, which would mean that the initial population density function would correspond to that of the seed crystals:

$$n(L, 0) = n_s(L) \tag{69}$$

where $\bar{n}_s$ is the population density function of the seed crystals.

Moments of the population density function, which are given by

$$m_j = \int_0^\infty L^j n dL \tag{70}$$

are especially useful in modeling crystal size distributions in batch operations and in the development of equations relating a control variable to time. Recognizing that the zero moment is the total number of crystals in the system it can be shown that

$$\frac{dm_0}{dt} = n^0 G - B^0 = \frac{d\bar{N}_T}{dt} \tag{71}$$

From the relationships of moments to properties of the distribution, the following equations can be derived:

$$\frac{dm_1}{dt} = Gm_0 + \frac{dL_T}{dt} = \bar{N}_T G \tag{72}$$

$$\frac{dm_2}{dt} = 2Gm_1 + \frac{dA_T}{dt} = 2k_a L_T G \tag{73}$$

$$\frac{dm_3}{dt} = 3Gm_2 + \frac{dM_T}{dt} = 3\left(\frac{k_v}{k_a}\right) \rho A_T G \tag{74}$$

where $\bar{N}_T$ is total number of crystals, $\bar{L}_T$ is total crystal length, $\bar{A}_T$ is total surface area of the crystals, and $\bar{M}_T$ is total mass of crystals in the crystallizer. A solute balance must also be satisfied:

$$\frac{d(V_C)}{dt} + \frac{dM_T}{dt} = 0 \tag{75}$$

where V is the system volume, and c is solute concentration in the solution.

Control of supersaturation is an important factor in obtaining crystal size distributions of desired characteristics, and it would be useful to have a model relating rate of cooling or evaporation or addition of diluent required to maintain a specified supersaturation in the crystallizer. Contrast this to the uncontrolled situation of natural cooling in which the heat transfer rate is given by

$$Q = UA(T - T_c)$$

(76)

where U is a heat-transfer coefficient, A is the area available for heat transfer, T is the temperature of the magma, and T_c is the temperature of the cooling fluid. If U and T_c are constants, the maximum heat transfer rate and the highest rate at which supersaturation is generated are at the beginning of the process when T is highest. These conditions can lead to excessive primary nucleation and the formation of incrustations on the heat-transfer surfaces.

Better product characteristics are obtained through control of the rate at which supersaturation (cooling, evaporation, and addition of a nonsolvent or precipitant) is generated. An objective of the operation may be to maintain the supersaturation at some constant prescribed value, usually below the metastable limit associated with primary nucleation. For example, the batch may be cooled slowly at the beginning of the cycle and more rapidly at the end.

Formulations of population balances on batch crystallizers have been illustrated, and a variety of operating strategies have been considered (65). The results are often complex and present difficult control schemes at best. For example, suppose a model is needed to guide the operation of a batch seeded crystallizer so that isothermal solvent evaporation can be accomplished at a rate that gives a constant crystal growth rate and no nucleation. It is shown that the evaporation rate required is a cubic function of time and the corresponding rate of heat input to the crystallizer must be controlled accordingly. If cooling was to be used rather than evaporation, a similar analysis would show that the dependence of crystallizer temperature on time is highly nonlinear. Although the development of a strategy for generating supersaturation can be aided by such analyses, the initial conditions in the models derived are based on properties of seed crystals added to the crystallizer.

The advantages of selective removal of fines from a batch crystallizer have been demonstrated (66,67). These experimental programs showed narrowing of crystal size distributions and suggest significant reductions in the fraction of a product that would consist of fines or undersize material.

Crystallizers and Crystallization Operations

Crystallization equipment can vary in sophistication from a simple stirred tank to a complicated multiphase column, and the operation can range from allowing a vat of liquor to cool through exchanging heat with the surroundings to the complex control required of batch cyclic operations. In principle, the objectives of these systems are all the same: to produce a pure product at a high yield with an acceptable crystal size distribution. However, the characteristics of the crystallizing system and desired properties of the product often dictate that a specific crystallizer be used in a particular operating mode.

Crystallization from Solution. Crystallization techniques are related to the methods used to induce a driving force for solids formation and to the medium from which crystals are obtained. Several approaches are defined in the following discussion.

Cooling crystallizers use a heat sink to remove both sensible heat from the feed stream and the heat of crystallization released as crystals are formed. The heat sink may be no more than the ambient surroundings of a batch crystallizer, or it may be cooling water or another process stream.

Evaporative crystallizers generate supersaturation by removing solvent, thereby increasing solute concentration. These crystallizers may be operated under vacuum, and, in such circumstances, it is necessary to have a vacuum pump or ejector as a part of the unit. If the boiling point elevation of the system is low (that is, the difference between the boiling point of a solution in the crystallizer and the condensation temperature of pure solvent at the system pressure), mechanical recompression of the vapor obtained from solvent evaporation can be used to produce a heat source to drive the operation.

Evaporative-cooling crystallizers are fed with a liquor that is at a temperature above that in the crystallizer. As the feed is introduced to the crystallizer, which is at reduced pressure, solvent flashes, thereby concentrating the solute in the resulting solution and reducing the temperature of the magma. The mode of this operation can be degenerated to that of a simple cooling crystallizer by returning condensed solvent to the crystallizer body.

Salting-out crystallization operates through the addition of a nonsolvent to the magma in a crystallizer. The selection of the nonsolvent is based on the effect of the solvent on solubility, cost, properties that affect handling, interaction with product requirements, and ease of recovery. The effect of adding a nonsolvent can be quite complex as it increases the volume required for a given residence time and may produce a highly nonideal mixture of solvent, nonsolvent, and solute from which the solvent is difficult to separate.

Reactive crystallization addresses those operations in which a reaction occurs to produce a crystallizing solute. The concentration of the solute formed generally is greater than that corresponding to solubility. In a subset of systems, the solubility is nearly zero and, concomitantly, the supersaturation produced by reaction is large. These are often referred to as *precipitation* operations, and crystal size distributions from them contain a large fraction of fine crystals.

Supercritical fluid solvents are those formed by operating a system above the critical conditions of the solvent. Solubilities of many solutes in such fluids often is much greater than those found for the same solutes but with the fluid at subatmospheric conditions. Recently, there has been considerable interest in using supercritical fluids as solvents in the production of certain crystalline materials because of the special properties of the product crystals. Rapid expansion of a supercritical system rapidly reduces the solubility of a solute throughout the entire mixture. The resulting high supersaturation produces fine crystals of relatively uniform size. Moreover, the solvent poses no purification problems because it simply becomes a gas as the system conditions are reduced below critical.

Crystallizers. The basic requirements of a system involving crystallization from solution are as follows: (1) a means of generating supersaturation in a fashion commensurate with the requirements of producing a satisfactory crystal size distribution, (2) a vessel to provide sufficient residence time for crystals to grow to a desired size, and (3) mixing to provide a uniform environment for crystal growth. There are numerous manufacturers of crystallization equipment; in addition, many chemical companies design their own crystallizers based on expertise developed within their organizations. Rather than attempt to describe the variety of special crystallizers that can be found in the marketplace, this section provides a brief general survey of types of crystallizers that use the modes outlined above. Greater detail can be found in the literature (68,69).

The forced-circulation crystallizer is a simple unit designed to provide high heat-transfer coefficients in either an evaporative or a cooling mode. Figure 22 shows a schematic diagram of a forced-circulation crystallizer that withdraws a slurry from the crystallizer body and pumps it through a heat exchanger where heat may be either added to or removed from the slurry. Heat transferred to the circulating magma causes evaporation of solvent as the magma is returned to the crystallizer, whereas heat removal lowers the temperature of the circulating magma. Forced circulation is used to control circulation rates and velocities past the heat-transfer surfaces.

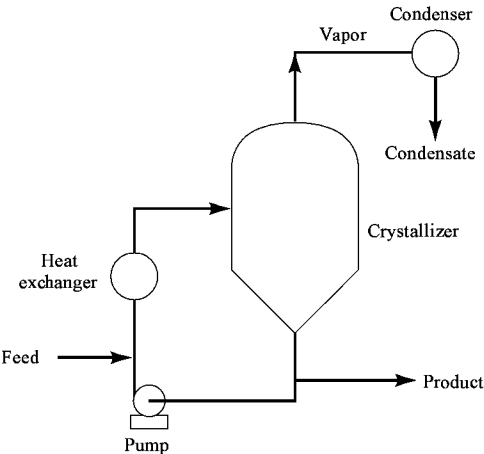


Fig. 22. Schematic diagram of forced circulation crystallizer.

When cooling is the selected mode by which supersaturation is generated, heat can be transferred through an external cooling surface, as shown in Figure 22, or through coils or a jacket internal to the crystallizer body. The higher heat-transfer coefficients that can be achieved with forced circulation allows the temperature difference between heat source and sink to be minimal, thereby reducing formation of encrustations on the heat transfer surface. The operation of cooling crystallizers is limited by the tendency of the solute to form encrustations on the cooling surface, so that the temperature of the cooling fluid and the temperature decrease of the slurry flowing through the heat exchanger may be limited. It is not uncommon to limit the decrease in magma temperature to about 3 to 5°C; therefore, both the circulation rate and heat-transfer surface must be large.

The feed in cooling crystallizers should be rapidly mixed with the magma so as to minimize the occurrence of regions of high supersaturation, which lead to excessive nucleation. Another factor that can lead to degradation of the crystal size distribution is the type of pump used in the circulation loop; an inappropriate pump can cause attrition of the crystals through abrasion, fracture, or shear, and most commercial systems use specially designed axial-flow pumps that provide high flow rates and low heads.

If the characteristics of the system are such that the operating temperature of the crystallizer is low in comparison to the temperature of cooling water, or if there are severe problems with the formation of encrustations, direct-contact refrigeration can be used. A refrigerant is mixed with the crystallizer contents and vaporized at the magma surface. On vaporizing, the refrigerant removes sufficient heat from the magma to cool the feed and remove the heat of crystallization. The refrigerant vapor must be compressed, condensed, and recycled for the process to be economical. Moreover, the refrigerant must be insoluble in the liquor to minimize losses and product contamination.

Scale formation on the heat exchanger surfaces or at the vapor-liquid interface in the crystallizer can cause operational problems with evaporative crystallizers. Such problems can be overcome by not allowing vaporization or excessive temperatures within the exchanger and by proper introduction of the circulating magma into the crystallizer. The latter may be accomplished by introducing the magma below the surface of the magma in the crystallizer, so that all vaporization occurs from a well-mixed zone or by introducing the magma so as to induce a swirling motion that is intended to dislodge encrustations from the wall of the crystallizer at the vapor-liquid interface.

Special devices for classification of crystals may be used in some applications. Figure 23 shows a draft-tube-baffle (DTB) crystallizer designed to provide preferential removal of both fines and classified product. Feed is introduced to the fines circulation line so that nuclei resulting from feed introduction can be dissolved as the stream flows through the fines dissolution exchanger. A quiescent zone is formed between the baffle extending into the chamber and the outside wall of the crystallizer. Flow through the quiescent zone can be adjusted so that crystals below a certain size (determined by settling velocity) are removed in the fines dissolution circuit.

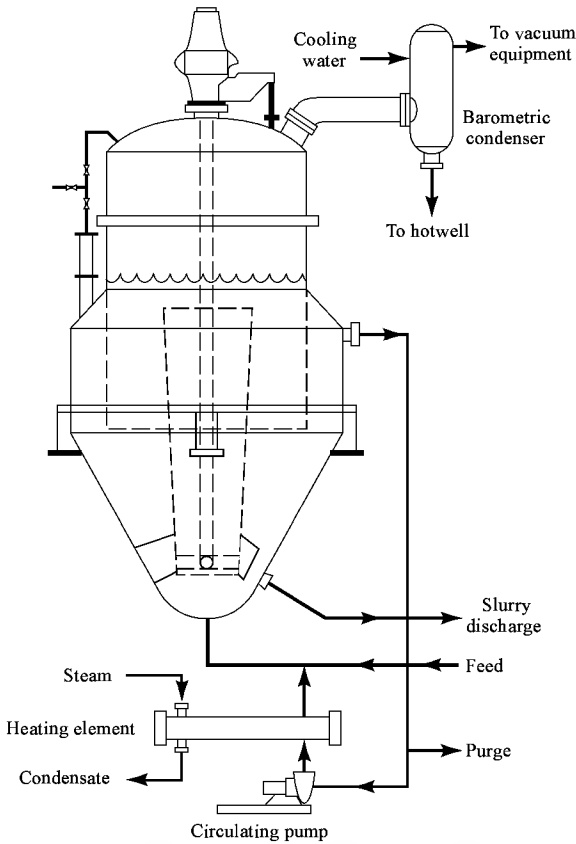


Fig. 23. Schematic diagram of draft-tube-baffle crystallizer.

Another type of crystallizer is the Oslo-type unit shown in Figure 24. In units of this type, the object is to form a supersaturated solution in the upper chamber and then relieve the supersaturation through growth in the lower chamber. The use of the downflow pipe in the crystallizer provides good mixing in the growth chamber.

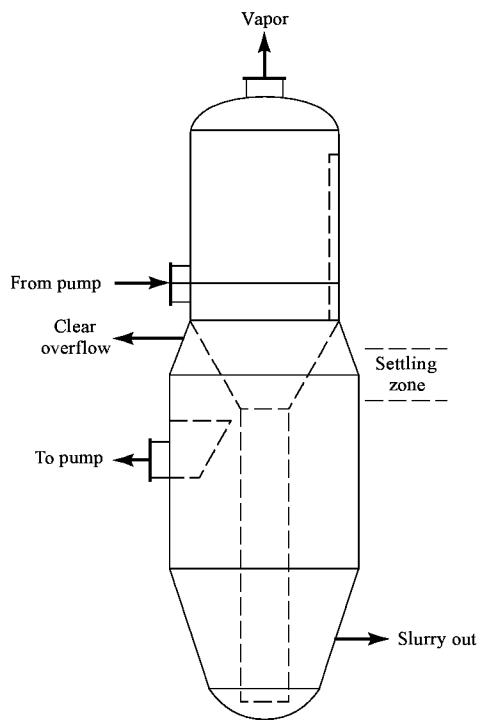


Fig. 24. Schematic diagram of an Oslo crystallizer.

Melt Crystallization. The use of a solvent can be avoided in some systems. In such cases, the system operates with heat as a separating agent, as do several processes involving crystallization from solution, but formation of crystalline material is from a melt of the crystallizing species rather than a solution.

For the following reasons, melt crystallization holds great promise in situations in which it can be substituted for crystallization from solution: (1) Without the need to recover and maintain the purity of a solvent, processing costs are reduced substantially. (2) Because there is no contaminated solvent to handle, melt crystallization may be more environmentally benign. (3) Energy costs found in evaporative crystallization obviously would be reduced if it is possible to produce a desired solid without the need to evaporate solvent. (4) Melt crystallization may be a reasonable alternative to other separation and purification processes, because the heat of vaporization of most volatile organic materials is between two and five times their heat of fusion. An analysis of the energy requirements in melt processes concludes that such processes can compete with other thermal separation techniques only if the plant is well designed and the process precisely controlled (70).

Melt crystallization is carried out either with a suspension of crystals or an advancing front (layer) of solids, although a more complete categorization of melt crystallization is available (71). Following is a brief review of processes in which melt crystallization is used; a more complete review, including a worked out case study for system design, is available (69).

A suspension of crystals formed from the melt may be contacted by well-mixed mother liquor or the crystals may be moved countercurrently to liquor flow in a vertical or horizontal column. In column crystallizers, crystals are moved in a specific direction by gravity or rotating blades. The crystals are melted by the addition of heat when they reach a designated end of the crystallizer; a portion of the melt is removed as product and the remainder is returned to the system to flow countercurrently to and to wash the product crystals.

One of the early column crystallizers was that introduced for the separation of xylene isomers (see XYLENE and ETHYLBENZENE). In this unit, shown schematically in Figure 25, *p*-xylene crystals are formed in a scraped-surface chiller above the column and fed to the column. The crystals move downward counter-currently to impure liquid in the upper portion of the column and melted *p*-xylene in the lower part of the column. Impure liquor is withdrawn from an appropriate point near the top of the column of crystals while pure product, *p*-xylene, is removed from the bottom of the column. The pulse unit drives melt up the column as reflux and into a product receiver.

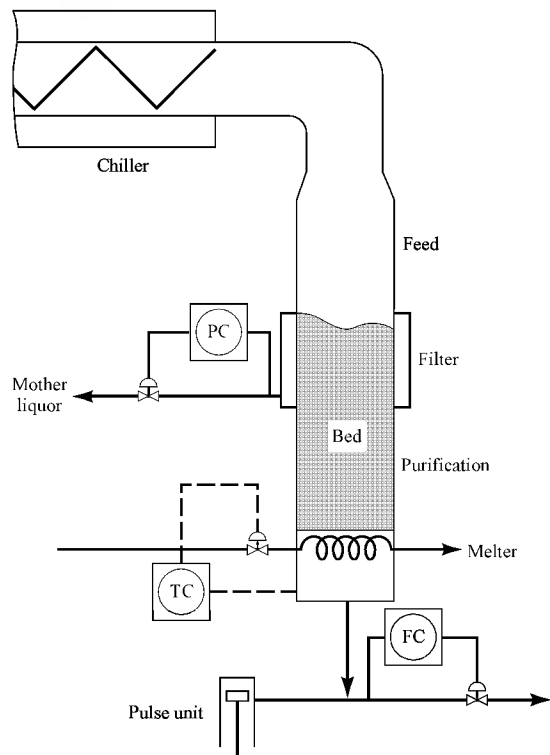


Fig. 25. Schematic diagram of a system used to separate xylene isomers (69). PC = pressure control, TC = temperature control, and FC = flow control.

A horizontal column is typified by the Brodie Purifier, which is shown schematically in Figure 26. Feed enters the column between recovery and refining sections, and crystals exit the refining section and pass through a purifying section. The purifying section is a wash column in which the crystals are contacted with melt generated at the bottom of the column.

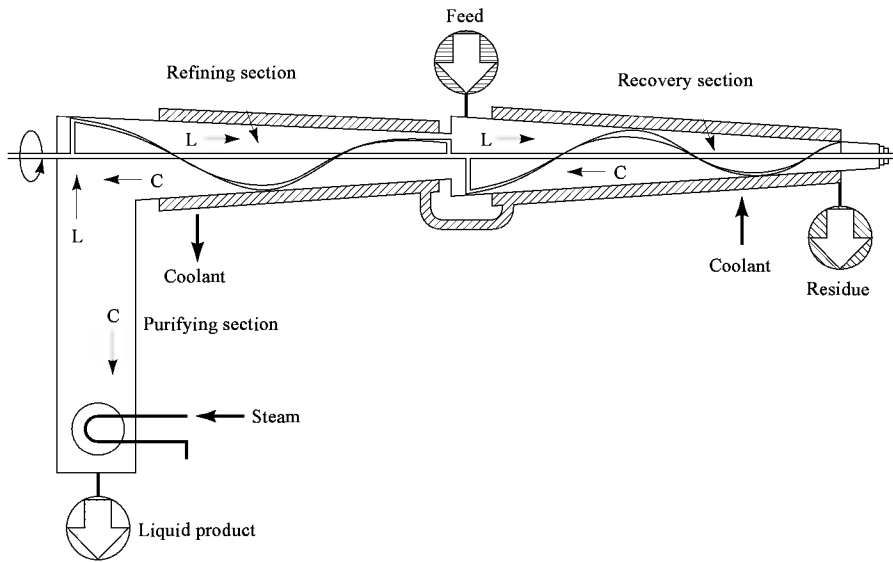


Fig. 26. Schematic diagram of Brodie Purifier. L = liquid; C = crystals (69).

In advancing-front or layer melt crystallizations, mother liquor flows over a cooled surface on which material is crystallized. The advancing front of crystals grows in the direction from the cooled surface into the mother liquor. A variety of techniques can be used to take advantage of this type of operation.

Figure 27 is a schematic diagram of the MWB process, which uses an operation in which there are several steps in a batch cycle. Crystal growth is on the inside of a battery of tubes through which melt is flowing, and the melt may flow as a falling film or it can be pumped through the tubes. The process includes the following steps:

- (1) Flow of mother liquor through the cooled tubes is initiated, and crystals are grown on the tube surfaces. The heat transfer rate should be controlled so as to moderate crystal growth, thereby producing a relatively uniform layer of high purity solids.
- (2) When sufficient crystal mass has been formed, melt flow is stopped and residual mother liquor is drained from the unit.
- (3) The solid purity is enhanced by applying heat and causing impurities to flow from the heated solids through a process known as sweating.
- (4) After sweating, either the cycle is returned to step 1 and additional solids are deposited or the solids present on the tube wall are melted and recovered as product.
- (5) The recovered product melt can be put through the cycle again to increase purity, or fresh feed can be introduced to the cycle.

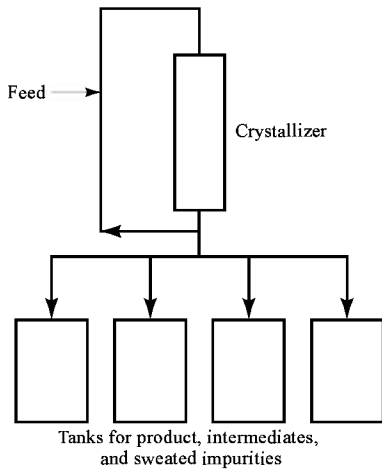


Fig. 27. Schematic diagram of the MWB melt crystallization process.

BIBLIOGRAPHY

"Crystallization," in *ECT* 1st ed., Vol. 4, pp. 619–636, by A. Ralph Thompson, University of Pennsylvania; in *ECT* 2nd ed., Vol. 6, pp. 482–515, by J. W. Mullin, University of London; in *ECT* 3rd ed., Vol. 7, pp. 243–285, by J. W. Mullin, University of London.

1. National Research Council, *Frontiers in Chemical Engineering: Research Needs and Opportunities*, National Academy Press, Washington, D.C., 1988, Chapt. 3.
2. R. M. Felder and R. W. Rousseau, *Elementary Principles of Chemical Processes*, 2nd ed., John Wiley & Sons Inc., New York, 1986.
3. R. H. Perry and D. W. Green, eds., *Perry's Chemical Engineers' Handbook*, McGraw-Hill Book Co., Inc., New York, 1984, p. 3–99.
4. H. Charmolue and R. W. Rousseau, *AIChE J.* **37**, 1121 (1991).
5. J. W. Mullin, *Crystallization*, 2nd ed., CRC Press, Cleveland, Ohio, 1972.
6. H. A. Miers and F. Issac, in H. E. Buckley, *Crystal Growth*, John Wiley & Sons, Inc., New York, 1951, p. 7.
7. A. D. Randolph and M. A. Larson, *Theory of Particulate Processes*, 2nd ed., Academic Press, Inc., New York, 1988.

8. J. Garside and Roger J. Davey, *Chem. Eng. Commun.* **4**, 393 (1980).

9. M. A. Larson, *AIChE Symp. Ser.*, **80**(240), 39 (1984).

10. M. W. Girolami and R. W. Rousseau, *Ind. Eng. Chem. Process Des. Dev.* **25**, 66 (1986).

11. C. Y. Sung, J. Estrin, and G. R. Youngquist, *AIChE J.* **19**, 957 (1973).

12. R. E. A. Mason and R. F. Strickland-Constable, *Trans. Faraday Soc.*, **62**(pt. 2), 455 (1966).

13. D. P. Lal, R. E. A. Mason, and R. F. Strickland-Constable, *J. Cryst. Growth* **5**, 1 (1969).

14. N. A. Clontz and W. L. McCabe, *Chem. Eng. Prog. Symp. Ser.* **67**(110), 6 (1971).

15. R. W. Rousseau and W. L. McCabe, *Paper presented at the World Congress on Chemical Engineering*, Amsterdam, June 1976.

16. R. C. Bennett, H. Fiedelman, and A. D. Randolph, *Chem. Eng. Prog.* **69**(7), 86(1973).

17. P. A. M. Grootcholten, L. D. M. v.d. Brekel, and E. J. deJong, *Chem. Eng. Res. Des.* **62**, 179 (1984).

18. J. Garside and M. B. Shah, *Ind. Eng. Chem. Process Des. Dev.* **19**, 509 (1980).

19. C. Y. Tai, W. L. McCabe, and R. W. Rousseau, *AIChE J.* **21**, 351 (1975).

20. C. Y. Tai, J.-F. Wu, and R. W. Rousseau, *J. Cryst. Growth* **116**, 294 (1992).

21. J. W. Mullin and C. L. Leci, *Phil. Mag.* **19**, 1075 (1969).

22. P. M. McMahon, K. A. Berglund, and M. A. Larson in S. J. Jančić and E. J. deJong, eds., *Proceedings of the 9th Symposium on Industrial Crystallization*, Elsevier, Amsterdam, The Netherlands, 1984, p. 229.

23. R. M. Ginde and A. S. Myerson, *J. Cryst. Growth* **116**, 41 (1992).

24. J. Garside, *AIChE Symp. Ser. No. 240*, **80**, 23 (1984).

25. R. F. Strickland-Constable, *Kinetics and Mechanism of Crystallization*, Academic Press, Inc., New York, 1968.

26. M. Ohara and R. C. Reid, *Modeling Crystal Growth Rates from Solution*, Prentice-Hall Press, Englewood Cliffs, N.J., 1973.

27. W. K. Burton, N. Cabrera, and F. C. Frank, *Philos. Trans. R. Soc. Lond.* **243**(A866), 299 (1951).

28. J. R. Bourne, *AIChE Symp. Ser. No. 193*, **76**, 59 (1980).

29. W. L. McCabe, *Ind. Eng. Chem.* **21**, 30 (1929).

30. E. T. White and P. G. Wright, *CEP Symp. Ser.* **67**(110), 81 (1971).

31. J. Garside and S. J. Jančić, *Chem. Eng. Sci.* **33**, 1623 (1978).

32. A. H. Janse and E. J. de Jong in E. J. de Jong and S. J. Jančić, eds., *Industrial Crystallization '78*, North-Holland, Amsterdam, The Netherlands, 1979, p. 135.

33. A. H. Janse and E. J. de Jong in J. W. Mullin, ed., *Industrial Crystallization*, Plenum Press, New York, 1976, p. 145.

34. M. W. Girolami and R. W. Rousseau, *AIChE J.* **31**, 1821 (1985).

35. K. A. Berglund and M. A. Larson, *AIChE Symp. Ser. No. 215*, **78**, 9 (1982).

36. A. D. Randolph and E. T. White, *Chem. Eng. Sci.* **32**, 1067 (1977).

37. C. Y. Lui, H. S. Tsuei, and G. R. Youngquist, *Chem. Eng. Prog. Symp. Ser.* **67**(110), 43 (1971).

38. H. J. Human, W. J. P. van Enckevork, and P. Bennema in J. J. Jančić and E. J. de Jong, eds., *Industrial Crystallization 81*, North-Holland, Amsterdam, The Netherlands, 1982, p. 387.

39. K. A. Berglund, E. L. Kaufman, and M. A. Larson, *AIChE J.* **25**, 867 (1983).

40. K. E. Blem and K. A. Ramanarayanan, *AIChE J.* **33**, 677 (1987).

41. R. C. Zumstein and R. W. Rousseau, *AIChE J.* **33**, 121 (1987).

42. N. A. Clontz, R. T. Johnson, W. L. McCabe, and R. W. Rousseau, *Ind. Eng. Chem. Fundam.* **11**, 368 (1972).

43. R. C. Zumstein and R. W. Rousseau, *Ind. Eng. Chem. Res.* **28**, 334 (1989).

44. S. N. Black, R. J. Davey, and M. Halcrow, *J. Cryst. Growth* **79**, 765 (1986).

45. M. Lahav and L. Leiserowitz in A. Mersmann, ed., *Proceedings of the 11th Symposium on Industrial Crystallization*, 1990, p. 609.

46. G. Henning in Ref. 45, p. 113.

47. M. J. Hounslow, R. L. Ryall, and V. R. Marshall, *AIChE J.* **34**, 1821 (1988).

48. P. Marchal, R. David, J. P. Klein, and J. Villermaux, *Chem. Eng. Sci.* **43**, 59 (1988).

49. R. C. Zumstein and R. W. Rousseau, *Chem. Eng. Sci.* **44**, 2149 (1989).

50. M. W. Girolami and R. W. Rousseau, *J. Cryst. Growth* **71**, 220 (1985).

51. N. Hiquily and C. Laguërie in S. J. Jančić and E. J. deJong, eds., *Industrial Crystallization '84*, Elsevier, Amsterdam, The Netherlands, 1984, p. 79.

52. C. Balarew in Ref. 38.

53. S. N. Black and R. J. Davey, *J. Cryst. Growth* **90**, 136 (1988).

54. K. G. Denbigh and E. T. White, *Chem. Eng. Sci.* **21**, 739 (1966).

55. D. D. Edie and D. J. Kirwan, *Ind. Eng. Chem. Fundam.* **12**, 100 (1973).

56. A. S. Myerson and D. J. Kirwan, *Ind. Eng. Chem. Fundam.* **4**, 414 (1977).

57. A. S. Myerson and D. J. Kirwan, *Ind. Eng. Chem. Fundam.* **4**, 420 (1977).

58. N. Hiquily, C. Laguërie, and J. P. Couderc, *Chem. Eng. J.* **30**, 1 (1985).

59. D. H. Jenkins and H. N. Sinha in Ref. 51.

60. R. C. Zumstein, T. Gambrel, and R. W. Rousseau in A. S. Myerson and K. Toyokura, eds., *Crystallization as a Separations Process*, ACS Symp. Ser. No. **438**, 1990, Washington, D.C., p. 85.

61. R. C. Zumstein and R. W. Rousseau, *AIChE Symp. Ser.* **83** (253), 130 (1987).

62. R. W. Rousseau and R. M. Parks, *Ind. Eng. Chem. Fundam.* **20**, 71 (1981).

63. J. R. Beckman and A. D. Randolph, *AIChE J.* **23**, 510 (1977).

64. R. W. Rousseau and T. R. Howell, *Ind. Eng. Chem. Process Des. Dev.* **21**, 606 (1982).

65. N. S. Tavare, J. Garside, and M. R. Chivate, *Ind. Eng. Chem. Process Des. Dev.* **19**, 653 (1980).

66. A. G. Jones, A. Chianese, and J. W. Mullin, in Ref. 51.

67. G. L. Zipp and A. D. Randolph in J. Nývlt and S. Žóček, eds., *Proceedings of the 10th Symposium on Industrial Crystallization*, Elsevier, New York, 1989, p. 469.

68. R. Bennett in Ref. 3, Chapt. 19.

69. R. W. Rousseau and C. G. Moyers in R. W. Rousseau, ed., *Handbook of Separation Process Technology*, John Wiley & Sons, Inc., New York, 1987, Chapt. 11.

70. K. Wintermantel and G. Wellinhoff in Ref. 45, p. 703.

71. J. Ulrich, Y. Iizoguz, and K. Wangnick in B. Biscans, N. Gabas, and C. Laguërie, eds., *Proceedings Crystallization Industrielle et Precipitation*, Lavoisier Technique et Documentation, Paris, 1991.

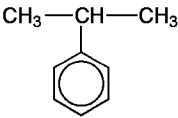
Ronald W. Rousseau
Georgia Institute of Technology

CRYSTALS.

See CRYSTALLIZATION; X-RAY TECHNOLOGY; ZONE REFINING.

CUMENE

The cumene molecule can be visualized as a straight-chain propyl group having a benzene ring attached at the middle carbon.



Thus cumene [98-82-8] (1-methylethylbenzene, 2-phenylpropane, isopropylbenzene), C₉H₁₂, is a substituted aromatic compound in the benzene (qv), toluene (qv), and ethylbenzene series (see BTX PROCESSING; XYLENES AND ETHYLBENZENE). Cumene is a clear liquid at ambient conditions. High purity cumene is normally manufactured from propylene and benzene and is a minor constituent of most gasolines. It is the principal chemical used in the worldwide production of phenol and its coproduct acetone. Many consumer or industrial products, such as plywood and composition board bonded with phenolic resins, nylon-6, epoxy and polycarbonate resins, and solvents, have origins that can be traced to cumene. Estimated U.S. production data for 1977 and 1987 were 1.2 and 1.8 million metric tons, respectively (1).

Properties

Some physical, chemical, and thermodynamic properties of cumene are listed in Tables 1 and 2 (2–6). Useful health and safety data have been included. The transport and other properties of cumene and other compounds in the series have been given in graphical form (6).

Table 1. Physical Properties of Cumene

Property	Value
freezing point, °C	96.03
boiling point, °C	152.39
density, g cm ³	
0°C	0.8786
20°C	0.8619
40°C	0.8450
refractive index, n _D ²⁰	1.4915
thermal conductivity at 25°C, W (m·K)	0.124
viscosity, mPas(cP)	
0°C	1.076
20°C	0.791
40°C	0.612
surface tension at 20°C, mN m (dyn cm)	28.2
vapor pressure, ^a kPa ^b	
35°C	1
100°C	19
120°C	37
140°C	68
180°C	185
flash point, °C	44
autoignition temperature, °C	523
flammable limits in air, vol %	
lower	0.9
upper	6.5

^a Calculated from the equation: ln *P* = *A* − *B* / (*t* + *C*); where *t* = temp, °C; *P* = vapor pressure, kPa^b; *A* = 13.99; *B* = 3400; and *C* = 207.78.

^b To converted kPa to mm Hg, multiply by 7.5.

Table 2. Chemical and Thermodynamic Properties of Cumene

Property	Value
relative molar mass	120.2
critical temperature, °C	351.4
critical pressure, kPa ^a	3220
critical density, g cm ³	0.280
heat of vaporization at bp, J g ^b	312

heat of vaporization at 25°C, J g ^b	367
heat of formation (liquid) at 25°C, J mol ^b	44,150
free energy (vapor) at 25°C, J mol ^b	137,000
heat of combustion at constant pressure and 25°C, J g ^b	
gross (product water as liquid)	43,370
net (product water as vapor)	41,170
heat capacity (liquid) at 25°C, J (molK) ^b	197
heat capacity (ideal vapor) at 25°C, J (molK) ^b	153
characteristic odor	aromatic
odor threshold, ppmv	1.2
threshold limit value, ppmv	50

^a To convert kPa to atm, divide by 101.3.
^b To convert from J to cal, divide by 4.184.

Manufacture

Cumene as a pure chemical intermediate is produced in modified Friedel-Crafts reaction processes that use acidic catalysts to alkylate benzene with propylene (see ALKYLATION; FRIEDEL CRAFTS REACTIONS). The majority of cumene is manufactured with a solid phosphoric acid catalyst (7). The remainder is made with aluminum chloride catalyst (8).

The most common process for making cumene is illustrated in Figure 1 (see also ALKYPHENOLS). The process uses an adiabatic reactor for the exothermic alkylation. A significant portion of the heat of reaction is recovered in preheating the feed and rectifying the effluent to generate a portion of the benzene recycle. Further cumene product purification includes recovery of the remaining benzene, clay treatment, and fractionation to remove small amounts of olefin oligomers and heavy material, respectively. The propylene feed may either be pure or contain a substantial amount of propane, which can come from a refinery fluid catalytic cracking operation. However, the feed must be essentially free of ethylene and butylenes to avoid contamination of the product with ethyl- and butylbenzenes.

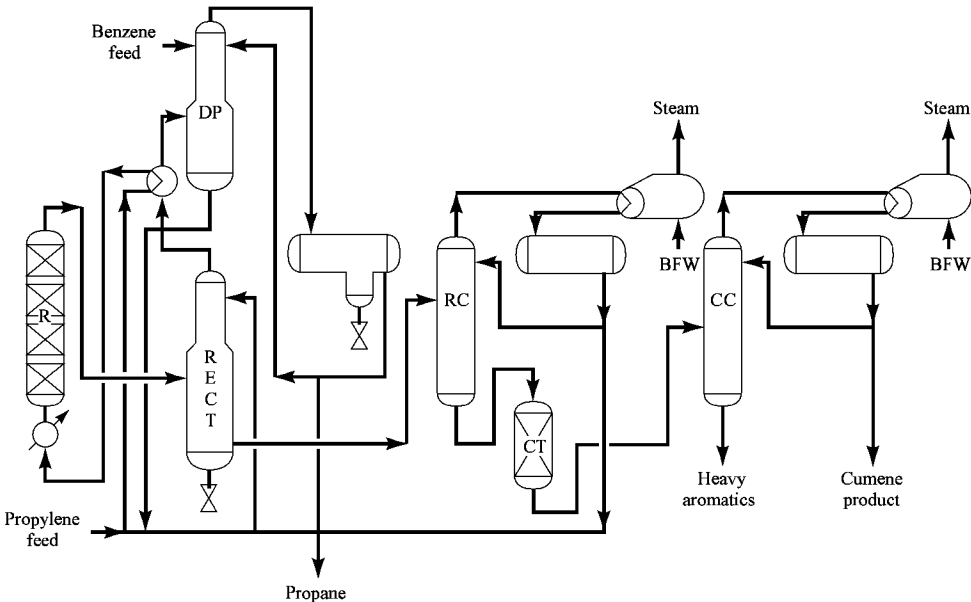


Fig. 1. UOP catalytic condensation process for cumene synthesis. R = reactor; RECT = rectifier; DP = depropanizer; RC = recycle column; CT = clay treater; CC = cumene column; and BFW = boiler feed water.

Courtesy of UOP.

Several other catalyst systems have been suggested, including boron fluoride and both crystalline and noncrystalline silicas and aluminosilicates. Although no commercial facility exists, the concept of using a crystalline silica or aluminosilicate catalyst in an integral reaction and distillation apparatus has been proposed (9).

Economic Aspects

Cumene production follows the demand for phenol and its derivatives. In 1987, the U.S. cumene demand, estimated at 1.9 million metric tons, was supplied by about 1.8 million metric tons of domestic production; imports made up the balance. Based on a trend from 1985 on, the demand is projected to increase at about 3% per year through 1995.

Costs per metric ton for cumene are difficult to assess because nearly half of the cumene is made captively; that is, the cumene is produced at a site, then further processed to phenol at the same site. Therefore, the "price" for the cumene remains secret.

From 1980 through 1987, the market price fluctuated between \$395 and \$595 per metric ton, ending up at \$550 per ton in 1987 (1,10). Since the beginning of 1990, the U.S. contract price for bulk quantities of cumene has shown a general downward trend. A compilation of prices from then until the end of April 1992 shows prices starting the period at \$520 per metric ton. For a brief period from late 1990 to early 1991, the price rose to levels around \$550 per ton, but from then until April 1992, the price decreased to a steady value of \$425 per ton (11).

In a modern operation the cost of production is made up of the factors shown in Table 3. The feedstocks account for more than 90% of the cost. Other operating costs, such as utilities, labor, and capital charges, make up the difference.

Table 3. Cost of Cumene Production per Metric Ton Cumene

Factor	Consumption, t	Typical cost, \$	Manufacturing cost, ¢
benzene charge	0.67	268	61
propylene charge, 95% purity	0.39	136.5	31

by-product (credit at \$250 /t)	(0.6)	(15)	(3)
utilities			
heat and electricity	1843 MJ ^a	5.2	1
steam credit	(0.6)	(4.2)	(1)
catalyst and chemicals		4.3	1
labor, insurance, maintenance, etc		7.6	2
interest and depreciation at 30% of capital		36.6	8
<i>Total</i>		<i>439</i>	

^a To convert MJ /t cumene to Btu /lb, divide by 2.324.

Specifications and Analytical Methods

Cumene sold as merchant grade for chemical purposes is usually produced to the specifications listed in Table 4. Captively manufactured cumene typically is not held to such strict values, although 99.9% purity is common. The composition specification does not include the small amounts, corresponding to about 400 ppm by weight, of olefinic material normally present that is indicated by the bromine index.

Table 4. Frequently Used Specifications for Cumene

Property	Test	Value
specific gravity, 15.5°C /15.5°C	ASTM D891	0.864 min /0.867 max
bromine index	ASTM D1492	50 max
color	ASTM D1209	10 max
sulfur compounds, as ppmw S	ASTM D4045	1 max
product assay		
cumene, wt %	ASTM D3760	99.9 min
ethylbenzene, ppm	ASTM D3760	200 max
n-propylbenzene, ppm	ASTM D3760	300 max
butylbenzenes, ppm	ASTM D3760	200 max

Health and Safety Factors

Cumene is a significant fire hazard when exposed to flame or sparks and is in the class of liquids that can be ignited under almost all normal temperature conditions (5). Cumene fires should be extinguished with foam, carbon dioxide, or dry chemical. Water may be ineffective. Because cumene vapors are heavier than air and may travel long distances, they could encounter an ignition source. When the concentration of cumene in air reaches about 0.1 ppm, its gasolinelike odor becomes apparent. An atmosphere containing up to 500 ppm cumene may be safely entered if a chemical cartridge respirator with an organic vapor cartridge is used. Personnel should not enter atmospheres containing 500 to 8000 ppm unless they wear Type C supplied-air respirators with full face piece. Wearing of contact lenses should be avoided any time cumene vapor has become apparent.

Cumene is a primary skin and eye irritant. Exposure may result in significant narcosis, headache, and nausea. Because the depressant action has a slow induction period and a long elimination period, possible cumulative effects need to be considered. The recommended threshold limit value (TLV) is 50 ppm (243 mg /m³), which is an 8-h time-weighted average for exposure to cumene (12). In the absence of human data, this value is recommended to prevent the induction of narcosis. The TLV contains a notation that cumene can be absorbed through the skin, and therefore this route should be considered when evaluating total exposure to cumene. The permissible exposure limit for cumene, given by the Occupational Safety and Health Administration, is also 50 ppm (245 mg /m³), again with a skin notation.

Cumene is expected to exist almost entirely in the vapor phase in the atmosphere (13). In water, mixed cultures of microorganisms collected from various locations and depths in the Atlantic Ocean were all found to be capable of degrading cumene (14). A number of studies have examined the aerobic degradation of cumene in seawater and in groundwater (15,16). The results indicate that cumene would normally be naturally degraded to below detectable limits within a week to ten days. Cumene is tightly adsorbed by soil and is not significantly mobile in soil (17).

A relatively small number of studies have reported on the effects of cumene on plants, fish, and other organisms. Studies of the effects of cumene on fresh and saltwater fish indicate the lowest reported toxic concentration (LC₅₀) for fishes was 20 to 30 mg /L (18). The solubility of cumene is about 50 mg /L (19). Among invertebrates, the lowest reported concentration that was toxic to test organisms was 0.012 mg /L after 18 hours (20). The only available data on the effect of cumene on aquatic plants indicate that the photosynthesis of several species was inhibited at concentrations from 9 to 21 mg /L (19).

Uses

More than 95% of the cumene produced is used as feedstock for the production of phenol (qv) and its coproduct acetone (qv). The cumene oxidation process for phenol synthesis has been growing in popularity since the 1960s and is prominent today. The first step of this process is the formation of cumene hydroperoxide [80-15-9]. The hydroperoxide is then selectively cleaved to phenol [108-95-2] and acetone [67-64-1] in an acidic environment (21).

Phenol, in its various purity grades, is used for phenol–formaldehyde resins to bond construction materials like plywood and composition board (40% of the phenol produced), for the bisphenol A employed in making epoxy resins (qv) and polycarbonate (qv) (30%), and for caprolactam (qv), the starting material for nylon-6 (20%). Minor amounts are used for alkylphenols (qv) and pharmaceuticals (10).

The acetone supply is strongly influenced by the production of phenol, and so the small difference between total demand and the acetone supplied by the cumene oxidation process is made up from other sources. The largest use for acetone is in solvents although increasing amounts are used to make bisphenol A [80-05-7] and methyl methacrylate [80-62-6]. α -Methylstyrene [98-83-9] is produced in controlled quantities from the cleavage of cumene hydroperoxide, or it can be made directly by the dehydrogenation of cumene. About 2% of the cumene produced in 1987 went to α -methylstyrene manufacture for use in poly(α -methylstyrene) and as an ingredient that imparts heat-resistant qualities to polystyrene plastics.

Cumene in minor amounts is used as a thinner for paints, enamels, and lacquers and to produce acetophenone, the chemical intermediate dicumylperoxide, and diisopropylbenzene. It is also a good solvent for fats and resins and, as such, has been suggested as a replacement for benzene in many of its industrial applications (22).

BIBLIOGRAPHY

"Cumene" in *ECT* 2nd ed., Vol. 6, pp. 543–546, by D. J. Ward, Universal Oil Products Co.; in *ECT* 3rd ed., Vol. 7, pp. 286–290, by D. J. Ward, UOP.
1. Z. Sedaghat-Pour, *Chemical Economics Handbook*, Stanford Research Institute International, Menlo Park, Calif., Mar. 1989.

2. R. W. Gallant, *Hydrocarbon Process* **48**(11), 263 (1969).

3. *Selected Values of Properties of Hydrocarbons*, Research Project 44, American Petroleum Institute, New York, 1953.

4. T. P. Thinh and co-workers, *Hydrocarbon Process.* **50**(1), 98 (1971).

5. *Cumene, Hazardous Substances Databank* No. 172, TOXNET (Toxicology Data Network), National Library of Medicine, Bethesda, Md., 1989.

6. C. L. Yaws, *Chem. Eng. (N.Y.)* **82**(20), 73 (1975).

7. R. C. Schulz and co-workers, "Cumene Technology Improvements," paper presented by UOP at the *AIChE Summer Meeting*, Denver, Colo., 1988.

8. R. C. Canfield and T. L. Unruh, *Chem. Eng. (N.Y.)* **90**(6), 32–33 (1983).

9. **U.S. Pat. 4,950,834** (Aug. 21, 1990), R. P. Arganbright and D. Hearn.

10. Chem Systems Inc., *Process Evaluation/Research Planning*, Report No. 85–10, 1986, p. 45.

11. *Chem. Mark. Rep.* **237–241** (Jan. 1990–Apr. 1992).

12. *Documentation of the Threshold Limit Values and Biological Exposure Indices*, Vol. 5, 5th ed., American Conference of Governmental Industrial Hygienists, Cincinnati, Ohio, 1986, p. 151.

13. S. J. Eisenreich and co-workers, *Environ. Sci. Technol.* **15**(1), 30–38 (1981).

14. J. D. Walker and co-workers, *Mar. Biol.* **34**(1), 1–9 (1976).

15. A. C. Van der Linden, *Dev. Biodeg. Hydrocarbons*, **1**, 165–200 (1978).

16. T. Kappelar and K. Wuhmann, *Water Res.* **12**, 327–342 (1978).

17. R. L. Swann and co-workers, *Res. Rev.* **85**, 17–28 (1983).

18. TSCA sec. 8(d) submission 878214937 *Summary of environmental data*, Dow item 1594, The Dow Chemical Company, Office of Toxic Substances, U.S. EPA, Washington, D.C., 1985.

19. T. C. Hutchinson and co-workers, *Environ. Sci. Res.* **16**, 577–586 (1980).

20. A. Rogerson and co-workers, *Aquat. Toxicol.* **3**(3), 215–228 (1983).

21. C. S. Hughes and Z. Sedaghat-Pour, in Ref. 1.

22. *Encyclopedia of Occupational Health and Safety*, Vols. I and II, Geneva, Switzerland, 1983, p. 572.

R. C. Schulz
P. J. Van Opdorp
D. J. Ward
UOP

CUMIN OIL.

See OILS, ESSENTIAL; FLAVORS AND SPICES.

CUPFERRON.

See ZINC.

CUPRITE.

See COPPER.

CURARE AND CURARELIKE DRUGS.

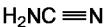
See ALKALOIDS.

CURIUM.

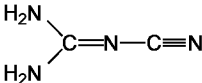
See ACTINIDES AND TRANSACTINIDES.

CYANAMIDES

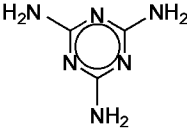
It has been suggested that under primordial conditions, cyanamide could have acted as the original peptide-forming and phosphorylating reagent at the beginning of life on earth (1). Calcium cyanamide [156-62-7], $\text{CaNC}\equiv\text{N}$, was first produced commercially around 1900 as fertilizer. A commercial process for manufacturing a pure grade of cyanamide was introduced in 1964. Later, numerous derivatives were developed, the most important of which are dicyandiamide (dimer) and melamine (trimer). However, its use in fertilizer has been replaced by ammonia and melamine is made principally by a petrochemical process. Structural formulas of cyanamide, CH_2N_2 (1), and its dimer $\text{C}_2\text{H}_4\text{N}_4$ (2), and trimer $\text{C}_3\text{H}_3\text{N}_3$ (3), are given as follows:



(1)



(2)



(3)

In North America, calcium cyanamide is no longer used as fertilizer, but it has limited use in special agricultural applications for defoliants, fungicides, herbicides, and as a weed killer. The primary industrial use is as a chemical intermediate for the manufacture of calcium cyanide, hydrogen cyanamide solution, and dicyandiamide. Calcium cyanamide is also used to add nitrogen to steel.

Cyanamide

Properties. Cyanamide [420-04-2], also called carbamodiimide or carbamic acid nitrile, crystallizes from a variety of solvents as somewhat unstable, colorless, orthorhombic, deliquescent crystals (2). Dimerization is prevented by traces of acidic stabilizers such as monosodium phosphate and by storage at low temperature. The ir spectrum of cyanamide exhibits an intense, poorly resolved doublet at $2225\text{--}2260\text{ cm}^{-1}$. The uv spectrum shows only weak absorption below 230 nm, with no maximum observable above 208 nm. Studies of the infrared and Raman spectra (3) support the *N*-cyanoamine structure, $\text{H}_2\text{N}-\text{C}\equiv\text{N}$. The properties of cyanamide are listed in Table 1.

Table 1. Properties of Cyanamide

Property	Value	References
molecular weight	42.04	
mp, °C	46	5
bp, °C		
101 kPa ^a	dec	
2.47 kPa ^a	140	
75 Pa ^a	85–87	
density ^b , g/mL	1.282	4
refractive index at 48°C	1.4418	
vapor pressure, up to 393 K	$\log \text{kPa}^a = \frac{-3.58 \times 10^3}{\text{K}} + 8.91$	4,6
infrared spectrum, cm ⁻¹	2225–2260	
dissociation constants		
K_a	5.4×10^{-11}	7
K_b	2.5×10^{-12}	8
specific heat at 0–39°C, J/(g K) ^c	2.288	9
heat of formation at 25°C, kJ/mol ^c	58.77	10
heat of solution ^d at 15°C, kJ/mol ^c	15.05	11
heat of combustion at 25°C, kJ/mol ^c	737.9	10
heat of neutralization ^e , kJ/mol ^c	15.48	12

^a To convert kPa to mm Hg, or Pa to μm Hg, multiply by 7.5.

^b Calculated.

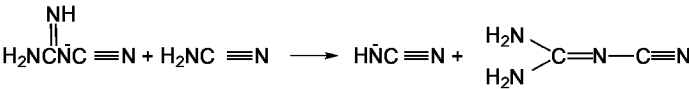
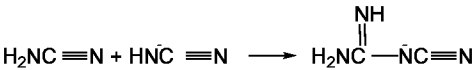
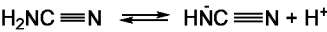
^c To convert J to cal, divide by 4.184.

^d In 1000 parts H₂O.

^e Excess NaOH solution.

Cyanamide is a weak acid with a very high solubility in water. It is completely soluble at 43°C, and has a minimum solubility (eutectic) at -15°C. It is highly soluble in polar organic solvents, such as the lower alcohols, esters, and ketones, and less soluble in nonpolar solvents (4).

Reactions. Reactions of cyanamide are either additions to the nitrile group or substitutions at the amino group. Both are involved in the dimerization to dicyandiamide.

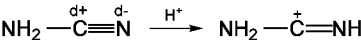


Dimerization involves addition of the cyanamide anion to the nitrile group of an undissociated molecule to give the anion of cyanoguanidine, or dicyandiamide. This reaction takes place most readily at pH 8–10 where the reactants are present in favorable proportion. The product is a weaker acid than cyanamide and is protonated at once with generation of a new cyanamide anion.

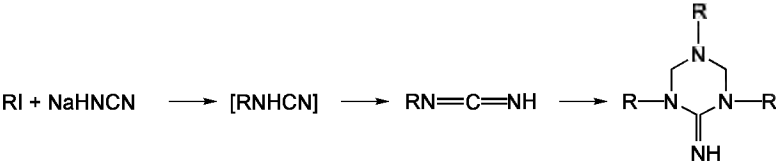
Most of the reactions occurring at the amino group of the cyanamide molecule require the anionic species, $\text{N}=\text{C}=\ddot{\text{N}}^-$ or $\text{HN}=\text{C}\equiv\text{N}$, sometimes in equivalent amount and occasionally as provided by base catalysis. Therefore, the process conditions for dimerization should be created to avoid the use of any metal salt, such as monosodium phosphate (4).

Similarly, nucleophilic reagents are suitable for addition reactions only if they are not so strongly basic as to produce the cyanamide anion in large amounts. In such cases, dicyandiamide is produced or a cyanamide salt is obtained. *N,N*-Disubstituted cyanamides do not ionize, of course, and react easily with strongly basic nucleophiles.

With nitriles in general, addition reactions take place readily in the presence of acids, because of the enhanced electrophilic character of the central carbon atom in the conjugate acid:



Substitutions. The cyanamide anion is strongly nucleophilic and reacts with most alkylating or acylating reagents (4); addition to a variety of unsaturated systems occurs readily (4). In some cases, a cyanamide salt is used; in others, base catalysis suffices. Ethyl iodide reacts with sodium hydrogen cyanamide [17292-62-5] to form a trisubstituted isomelamine.



Alkylation with a variety of common alkyl halides or sulfates gives stable dialkylcyanamides. However, the intermediate monoalkylated compounds usually cannot be isolated and cyclic trimers or cotrimers with cyanamide are obtained (13). The reaction can be carried out efficiently in water or alcohol. Allyl chloride is an especially useful reagent, producing diallylcyanamide [538-08-9] (4).



The amino group in cyanamide is very reactive toward formaldehyde. The reaction produces first an unstable hydroxymethyl derivative that

resinifies more or less rapidly, passing through a water-soluble stage which permits various applications. This product can be dried and thermoset. The reactions leading to resinification include condensation to give methylene and methyl ether bridges, hydration of the cyano group if acidic conditions are used, and polymerization of cyanamide residues to give dicyandiamide, melamine, or possibly isomelamine structures (see AMINO RESINS AND PLASTICS).

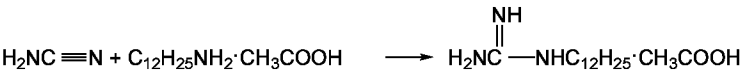
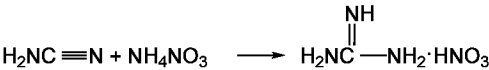
Cyanogen chloride and NaOH give the potentially useful sodium dicyanamide [1934-75-4] salt (14).



Other substitution reactions have been described with ketones, epoxides, anhydrides, acyl halides, amides, and imidates, among others (4).

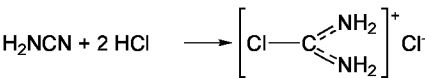
Additions. The addition reactions of ammonia and amines to the cyanamide nitrile group have been thoroughly studied (15). For optimum conditions, the reaction should be carried out in an aqueous medium at about 140°C. Gradual addition of the cyanamide to the amine salt minimizes dimerization.

A great variety of amino compounds have been used (4). The two examples given below have practical applications. In many cases it is advantageous to use anhydrous cyanamide in alcoholic solvents.

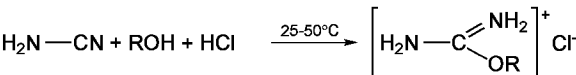


Uses of such guanidine derivatives depend on their strongly basic character and the cationic character of the salts. For example, 2-aminopyrimidine (guanidine derivative) is used in sulfa drugs, guanidine nitrate is used in the production of nitroguanidine (raw material for making explosives), and dodecylguanidine acetate is used as a fungicide (4).

Hydrogen chloride gives a dihydrochloride. Chloroformamidine hydrochloride [29671-92-9] is a convenient anhydrous form of cyanamide that is easily handled and stored (16).

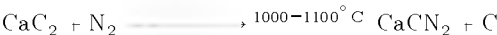


Addition of alcohols and phenols in the presence of anhydrous hydrogen chloride gives O-substituted pseudourea salts (17). The reaction is sluggish except with the lower alcohols, and long reaction time and temperatures up to 100°C are required to obtain good yields.



Manufacture.

From Limestone and Coal. The basic process for the manufacture of cyanamide comprises four stages. The first three steps produce calcium cyanamide: (1) lime is made from high grade limestone (see LIME AND LIMESTONE); (2) calcium carbide is manufactured from lime and coal or coke (see CARBIDES); (3) calcium cyanamide is produced by passing gaseous nitrogen through a bed of calcium carbide [715-20-7] with 1° calcium fluorspar, which is heated to 1000–1100°C in order to start the reaction. The heat source is then removed and the reaction continues because of its strong exothermic character.



The batch oven consists of a cylindrical refractory-lined steel shell with a steel cover lined with insulating material. A vent in the cover permits the escape of off-gases. The charge is either added directly to the oven or inserted in a perforation basket. The ovens are heated by carbon electrodes. Batch ovens range in size from 1–8 metric tons capacity. A cycle time of 4–5 days is required for 4-t ovens using a single starting electrode; with multiple electrodes, shorter times are possible. The reaction is almost completed in three days; the remaining time is required for cooling to avoid surface oxidation when the pig is withdrawn from the oven. A detailed account of the oven operation has been given (18).

A 1990 study by American Cyanamid Co. demonstrated that the reaction cycle depends also on the supply pressure of the gaseous nitrogen. The 3-d reaction cycle at the normal 75 Pa (0.3-in. water) pressure can be reduced to 2.4 d using the N₂ gas pressure at 2 kPa (8-in. water) into the ovens (19).

Continuous kilns were first used in Germany at Knapsack (20) and later at Trostberg (21). In the Knapsack process, calcium carbide (0.75–2 mm in size) is fed to a rotary kiln, 3 m in diameter and 12 m long with 1° slope; 1–2° calcium chloride is added to promote the reaction. The kiln produces 12–13 t fixed nitrogen per day. The product is granular and can be sold without further processing.

The Sdddeutsche Kalkstickstoffwerke process at Trostberg uses powdered calcium carbide along with recycle product and calcium fluoride in a rotary kiln at 1000–1100°C. The capacity of a unit is 25 t fixed nitrogen per day. The product passes to a rotary cooler and is granular (21).

For the manufacture of calcium cyanamide, crude calcium carbide (ca 3.36 × 1.68 mm or 6 × 12 mesh) can be used, whereas for cyanamide and dicyandiamide, a 74 μm (200 mesh) anhydrous carbide is used.

For agricultural uses, a fully hydrated calcium cyanamide powder containing 2° oil is used for dusting applications; for general fertilizer use, the hydrated material is granulated with water in rotary drums and dried giving a 2.38 × 0.32 mm (8 × 48 mesh) product.

In the final fourth step cyanamide is manufactured from calcium cyanamide by continuous carbonation in an aqueous medium (Fig. 1) (22).

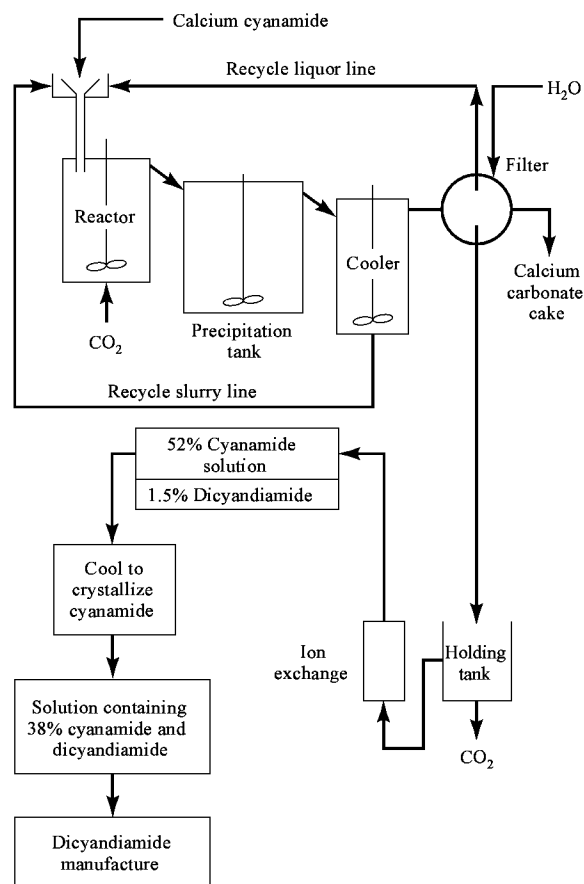
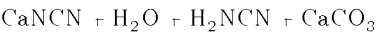


Fig. 1. Manufacture of cyanamide from calcium cyanamide.



A carbonated slurry of cyanamide solution, solid calcium carbonate, and graphite is cooled to remove the heat of reaction. Part of the slurry is recycled to facilitate temperature control whereas the remainder is filtered yielding cyanamide solution and a cake of calcium carbonate and graphite. The filtered solution is also recycled in order to control the solids content. The final concentration of cyanamide is normally maintained at 25°.

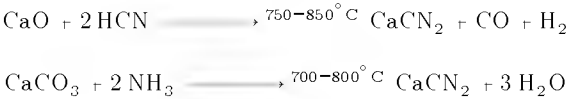
The calcium cyanamide feed is well mixed with the recycled slurry and filtrate in a feed vessel. The calcium cyanamide is added at a rate to maintain a pH of 6.0–6.5 in the cooling tank. The carbonation step can be conducted in a turbine absorber with a residence time of 1–2 min. After the carbonation step, the slurry is held at 30–40°C to complete the formation of calcium carbonate, after which the slurry is cooled and filtered. All equipment for the process is preferably of stainless steel. The resulting solution is used directly for conversion to dicyandiamide.

The filtered cake produced from the manufacture of dicyandiamide contains about 86° calcium carbonate. American Cyanamid Co. blends the dried waste for the manufacture of calcium carbide-based desulfurized reagents as a gas releasing agent.

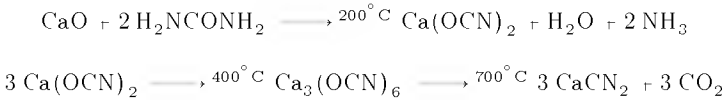
For production of commercial 50° solution and for recovery of crystalline cyanamide, this process is modified to improve purity and concentration. Calcium and iron may be removed by ion-exchange treatment. The commercial 50° solution is stabilized at pH 4.5–5.0 with 2° monosodium phosphate and contains less than 1.5° dicyandiamide and 0.2° urea. Such solutions are expected to show less than 1° change in cyanamide content per month of storage below 10°C. It is advisable, however, to adjust the pH periodically during extended storage. Organic esters may be used instead for improved stability (23).

Of the four steps described above, only the first two consume large amounts of energy, especially the calcium carbide step. With increasing energy costs, attention should be focused on alternatives for making calcium cyanamide.

Other Processes. Other methods include reaction of lime, (CaO) with hydrogen cyanide (24) and reaction of limestone, (CaCO₃), with ammonia (25).



Reaction of lime and urea forms calcium cyanate [6860-10-2] (26) which is then converted to calcium cyanurate [3266 5-90-0]; the latter gives calcium cyanamide at a higher temperature:



The product obtained is white calcium cyanamide whereas the product obtained from limestone and coal contains carbonaceous (graphite) impurities. None of these processes has been commercially exploited.

Economic Aspects. A peak in calcium cyanamide production was probably reached in 1962 when the world production for fertilizer use was of the order of 1,000,000 metric tons of calcium cyanamide per year, and for industrial use approximately 300,000 t (excluding the then USSR). In 1990, the total production of cyanamide products was about half that of 1962. The largest producers are in Japan, Germany, and Canada.

Altogether approximately forty plants are capable of producing calcium cyanamide. In North America, the only producer is Cyanamid Canada, Inc. in Niagara Falls, Ontario, serving the Canadian and United States markets. This company, a subsidiary of American Cyanamid Co., also produces dicyandiamide, cyanamide, and a number of their derivatives including calcium cyanide. Cyanamide is occasionally marketed in North America in crystalline form, but mostly as 50° solution. In Europe, cyanamide is available as 50° solution and in crystalline form from the S ddeutsche Kalkstickstoff-Werke A.G., Trostberg, Germany.

Specifications and Analysis. Cyanamide is sold as anhydrous, aqueous 50°, and calcium cyanamide. Aqueous 50° cyanamide solutions contain a buffer additive, usually 2° NaH₂PO₄, to stabilize the pH and prevent formation of dicyandiamide and urea. Calcium cyanamide is stable under dry conditions. Table 2 gives a typical analysis of the three commercial forms.

Table 2. Typical Analysis of Commercial Cyanamide^a

Assay	⁰ ₀
<i>Anhydrous Cyanamide-100</i>	
cyanamide	96
dicyandiamide	1.0
urea	0.5
water	0.5
NaH ₂ PO ₄ (stabilizer)	2.0
<i>Aqueous 50⁰ o Cyanamide-50</i>	
cyanamide	50
dicyandiamide	1.5
urea	0.2
water	46.3
NaH ₂ PO ₄ (stabilizer)	2
<i>Calcium cyanamide</i>	
calcium cyanamide	65
CaO	12
carbon	12.5
CaF ₂	0.5
CaS	1.0
calcium carbide	0.5
other nitrogen impurities	1.5
other mineral impurities	7

^a Ref. 4.

Cyanamide is precipitated by excess ammoniacal silver nitrate as disilver cyanamide [3384-87-0] which is dissolved in acid and titrated with thiocyanate solution (27).

Traces of cyanamide can be determined colorimetrically by complex formation with 1,2-naphthaquinone-4-sodium sulfonate (4).

Calcium cyanamide can be analyzed by determining the N₂ content using a combustion–conductivity method (28).

Handling and Storage. Cyanamide solution dimerizes to dicyandiamide and urea with the evolution of heat and a gradual increase in alkalinity accelerating the reaction. Storage above 30°C without pH stabilizer leads to excessive dimerization and can result in violent exothermic polymerization. Cyanamide should be stored under refrigeration and the pH tested periodically. Stabilized cyanamide can be kept at ambient temperature for a few weeks.

In general, cyanamide should be added to a reaction mixture at such a rate that it is used up as it is added, otherwise a high concentration of cyanamide results which could react violently.

Health and Safety Factors (Toxicology). Manufacture of cyanamide and calcium cyanamide does not present any serious health hazard. Ingestion of alcoholic beverages by workmen within several hours of leaving work sometimes results in a vasomotor reaction known as cyanamide flush. Cyanamide interferes with the oxidation of alcohol and accumulation of acetaldehyde probably accounts for this temporary phenomenon. Although extremely unpleasant, it has not been known to result in serious illness or to have any permanent effect.

Commercial grades of calcium cyanamide contain lime and are moderate skin irritants where contact is repeated or prolonged.

Contact or ingestion of cyanamide must be avoided, and precautions taken to prevent inhalation of dust or spray mist. In rat studies cyanamide-100 toxicity ranges from a single oral dose LD₅₀ of 280 mg kg to a single dermal dose LD₅₀ of 590 (420–820) mg kg. The compound is, therefore, considered to be moderately toxic both by ingestion in single doses and by single-skin applications. An aqueous paste of the product is corrosive to rabbit skin. Small quantities of the dry product produced severe irritation when introduced into the conjunctival sac of the rabbit eye.

Cyanamide-50 is considered to be slightly toxic by ingestion in single dose and moderately toxic by single-skin applications. The degree of irritation to rabbit skin and eye produced by this product is only slightly less than that observed with Cyanamide-100.

Uses. In Europe cyanamide and calcium cyanamide are used as fertilizers (qv), weed killers, and defoliants. In North America these applications have been practically discontinued. However, calcium cyanamide is used as a weed killer for asparagus and onions. Heavy applications approximately one month before planting are used to control soil-borne plant disease and weed seeds (see HERBICIDES). A direct application of calcium cyanamide is particularly useful to fertilize acid soils where lime values are needed. It can also be used in mixed fertilizers although the high alkalinity reduces the soluble phosphate content if an excess is used. It reduces severity of clubroot symptoms and improves yield in cabbage and brussel sprouts (29). It is a disinfectant for destroying ascaris eggs in infected pig manure (30). It is used to cure weed infestation against fungal disease in wheat, barley, and sugar beet (31). In California, some farmers use it as soil sterilant for tomato and celery disease and combating the potato eelworm. It has yielded excellent results in bud breaking in grape growing (32).

Calcium cyanamide has been used in the treatment of alcoholics. A small pill, taken once a day, subjects the user to cyanamide flush if any alcohol is taken. The result is unpleasant and discourages drinking (33).

Industrial uses make up most of the market for cyanamide. Calcium cyanamide is used directly for steel nitridation (34) and to some extent for desulfurization (36) (see STEEL). Cyanamide is used to produce cationic starch (36) and calcium cyanide. Cyanamide is, of course, the raw material for dicyandiamide and melamine. New uses include intermediates for pesticides, detergents (37), medicines such as antihistamines, hypertension, sedatives, contraceptives, etc (38), the photography industry (39), as an additive for fuels and lubricants, as a paper preservative, and as a cement additive.

Dicyandiamide

Properties. Dicyandiamide (2) (cyanoguanidine [461-58-5]) is the dimer of cyanamide and crystallizes in colorless monoclinic prisms. It is amphoteric, and generally soluble in polar solvents and insoluble in nonpolar solvents. Its properties are listed in Table 3.

Table 3. Properties of Dicyandiamide^a

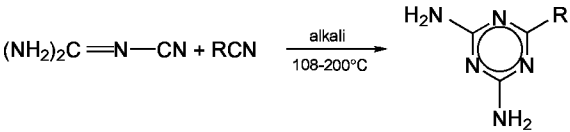
Property	Value
mol wt	84.08
mp, °C	208
bp, °C	dec
acid dissociation at 25°C, <i>K_a</i>	6 × 10 ⁻¹⁵
specific heat at 25°C, J (gK) ^b	1.41
heat of formation at 25°C, kJ mol ^b	24.9

heat of combustion at 25°C, kJ mol ^b	1382
heat of solution at 15°C, kJ mol ^b	24.1

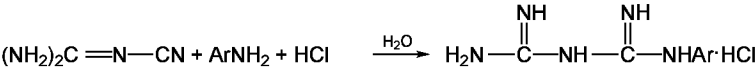
^a Refs. 40–42.

^b To convert J to cal, divide by 4.184.

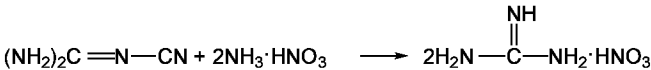
Reactions. The reactions of dicyandiamide resemble those of cyanamide. However, cyclizations take place easily and the nitrile group is less reactive. Under pressure and in the presence of ammonia, dicyandiamide cyclizes to melamine. Considerable tonnages of melamine have been made in this manner; however, melamine is produced chiefly by the urea process (43). Excellent yields of guanamines (44) are obtained when dicyandiamide is heated with alkyl or aryl nitriles in the presence of small amounts of alkali.



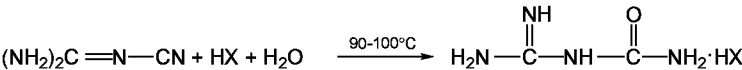
Heating with aromatic amines in water in the presence of an equivalent amount of mineral acid gives high yields of aryl biguanide salts (45).



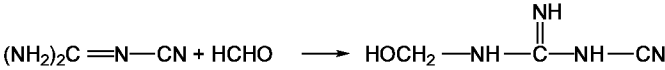
Reaction with ammonium salts gives biguanide salts which react further with the ammonium salt forming guanidine salts. Guanidine nitrate [506-93-4] is manufactured by this route (46).



Hydrolysis (47) of dicyandiamide occurs easily at elevated temperatures in the presence of an equivalent of mineral acid to yield guanylurea salts. This reaction is quantitative and can be used for the determination of dicyandiamide (48).



Dicyandiamide may be treated with formaldehyde (49) to produce resinous compositions of varying properties (see AMINO RESINS AND PLASTICS), under either acid or alkaline conditions. The reaction can be controlled to give mainly monomethyloldicyandiamide, a very water-soluble compound:



Manufacture. Dicyandiamide is manufactured by dimerization of cyanamide in aqueous solution. The 25° cyanamide solution produced as described above is adjusted to pH 8–9 and held at approximately 80°C for two hours to give complete conversion. The hot liquor is filtered and transferred to a vacuum crystallizer where it is cooled. The crystals of dicyandiamide are separated in continuous centrifuges and passed to rotary driers. The finished material is stored in bulk or in bags. The mother liquor is partly recycled to the cyanamide extraction system and partly purged to keep the thiourea content low. The process is environmentally safe. The process effluent discharges meet local government discharge guidelines. In 1990, the total worldwide production of dicyandiamide was about 30,000 tons.

Specifications, Analysis, and Toxicity. Dicyandiamide is identified qualitatively by paper chromatography and quantitatively by ultraviolet spectrometry of the chromatogram. More commonly, total nitrogen analysis is used as a purity control or the dicyandiamide is converted by hydrolysis to guanylurea, which is determined gravimetrically as the nickel salt (50). Methods based on the precipitation of silver dicyandiamide picrate are sometimes used (51). Dicyandiamide can also be titrated with tetrabutylammonium hydroxide in pyridine solution. Table 4 gives a typical analysis of a commercial sample. Dicyandiamide is essentially nontoxic. It may, however, cause dermatitis.

Table 4. Typical Analysis of Commercial Dicyandiamide

Assay	Value
dicyandiamide, %	99.3
water, %	0.01
melamine, %	0.7
thiourea, ppm	200
heavy metals, ppm	10

Uses. Dicyandiamide is used as a raw material for the manufacture of several chemicals, such as guanamines, biguanide and guanidine salts, and various resins. Since 1975, it has also been used in the manufacture of potassium or sodium dicyanamide which are used as insecticides and in chemotherapy. Melamine has extensive applications in the resin and plastic industry; guanamines are used as copolymers (qv) in many resin compositions. Guanidine phosphate [1763-07-1] is employed as a fire retardant in applications where water solubility is not a drawback.

Dicyandiamide, in conjunction with phosphoric acid, is used as a flame retardant for cellulosic materials (52). Cotton fabrics have also been treated effectively in this manner. Use as a fire retardant for wood, particularly shingles, has found commercial application (53).

Dicyandiamide reduces the viscosity of certain colloidal solutions. This property is of commercial significance in the manufacture of glues and adhesives, in the coating and sizing of paper and textiles, and in the conditioning of phosphate drilling muds (see PETROLEUM). This action may prove useful in other applications where control of viscosity is important (54).

Dicyandiamide fluidifies various adhesives and glues. In the presence of dicyandiamide, the initial viscosity of the adhesive formulations is lowered and maintained at a lower range for longer periods of time, thus preventing premature gelling and extending the useful life of the adhesive (see ADHESIVES).

Dicyandiamide, although completely stable at lower temperatures, has found use in surface coating formulations as a flame retardant. When exposed to a flame, the coating composition intumesces or bubbles up, thereby insulating the substrate. In this manner, the substrate retains its physical properties and strength. For example, one effective intumescent formulation includes 10 parts of dicyandiamide, 22 parts of pentaerythritol, 56 parts of monoammonium phosphate, 12 parts of titanium dioxide, 50 parts of latex binder (50° solids), and 33 parts of water. Such a formulation gives good foam volume at low coating weights, and reasonable stability of the foam structure under flaming (55) (see FOAMS).

Dicyandiamide has been marketed as a surface protective coating for printed circuit boards (56), in thermosetting resins (57), as an antistatic agent for cardboard and paper (58), for heat and light stabilization of PVC, retannage of various leathers (59), in improving adhesion of paints (60), as biocide, stabilizer of varnishes (61), and a curing agent for epoxy resins (62).

Melamine

Although melamine [108-78-1] (3) was prepared for the first time in 1834 by Liebig, almost a century passed before its properties and reactivity were thoroughly investigated. In the late 1930s a commercial process was developed. Since then, uses for melamine reaction products have been developed rapidly in a variety of fields, including plastics, surface coatings, bonding agents, paper and textile finishes, tanning agents, pharmaceuticals, and petroleum and rubber chemicals. Other names for melamine are cyanurotriamide, cyanuramide, and 2,4,6-triamine-1,3,5-triazine.

Properties. The outstanding characteristic of melamine, usually a white crystalline material, is its insolubility in most organic solvents. This property is also evident in melamine resins after they are cured. On the other hand, melamine is appreciably soluble in water, its solubility increasing with increased temperature. The properties of melamine are listed in Table 5.

Table 5. Properties of Melamine^a

Property	Value
molecular weight	126.13
mp, °C	350
density, g mL	1.573
specific heat at 25°C, J (gK) ^b	1.23
heat of combustion at 25°C, kJ mol ^b	1964
heat of sublimation, kJ mol ^b	121

^a Ref. 63.
^b To convert J to cal, divide by 4.184.

The chemistry of melamine has been reviewed (63,64). Melamine, although moderately basic, is better considered as the triamide of cyanuric acid than as an aromatic amine (see CYANURIC AND ISOCYANURIC ACIDS). Its reactivity is poor in nearly all reactions considered typical for amines. In part, this may be a result of its low solubility (see AMINO RESINS AND PLASTICS).

Manufacture. Dicyandiamide is converted into melamine by heating. Simple pyrolysis above the melting point leads to an exothermic reaction; however, deammoniation occurs, forming products containing two or three triazine rings as well as melamine. After it was discovered in 1940 that deammoniation can be counteracted by conducting the reaction under ammonia pressure, various methods were developed to control the exothermic reaction on an industrial scale.

By now the dehydration condensation of urea [57-13-6] has displaced the dicyandiamide process (see UREA). Although the latter is still used occasionally, the urea process predominates in North America. A flow sheet is shown in Figure 2 (43).

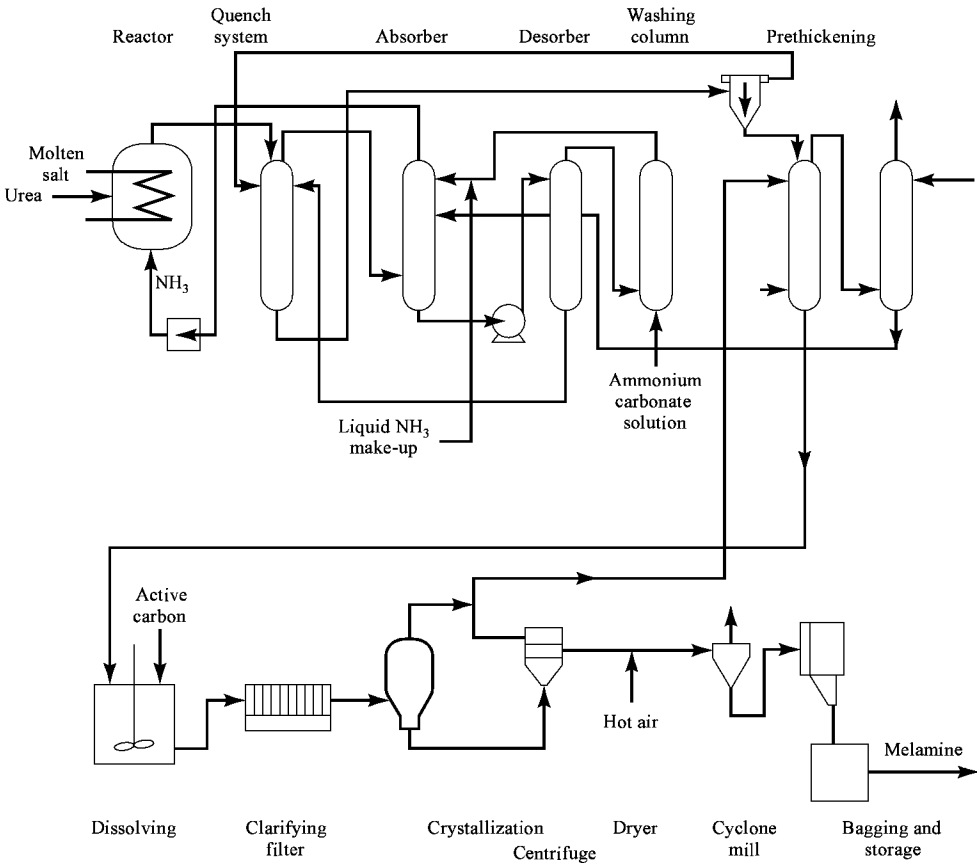
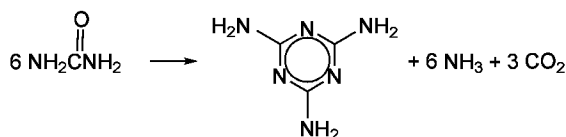


Fig. 2. Production of melamine for urea (43).

Courtesy of Gulf Publishing Co.

A urea melt is supplied to a one-stage reactor containing a fluid-bed catalyst. The reactor is heated internally by circulating molten salt. Upon entering the reactor, the urea is converted to melamine by the hot catalyst.



The hot ammonia used as carrier gas serves simultaneously as fluidizing agent and inhibits deammonization.

The reactor effluent is rapidly quenched with aqueous mother liquor in specially designed equipment operating at pressures essentially equal to the reactor pressure. This operation yields an off-gas consisting of ammonia and carbon dioxide vapor and a crystalline melamine slurry saturated with ammonia and carbon dioxide. The slurry is concentrated in a cyclone mill. The mother liquor overflow is returned to the quenching system. The concentrated slurry is redissolved in the mother liquor of the crystallization system, and the dissolved ammonia is stripped simultaneously.

This ammonia is recycled to the reactor via a compressor and a heater. Liquid ammonia is used as reflux on the top of the absorber. The net amount of carbon dioxide formed in the reactor is removed as bottom product from the absorber in the form of a weak ammonium carbamate solution, which is concentrated in a desorber-washing column system. The bottom product of this washing column is a concentrated ammonium carbamate solution which is reprocessed in a urea plant. The top product, pure ammonia, is liquefied and used as reflux together with liquid makeup ammonia. The desorber bottom product, practically pure water, is used in the quench system in addition to the recycled mother liquor.

In order to upgrade the melamine, the solution obtained after prethickening and stripping is treated with activated carbon passed over a clarifying filter and fed to a two-stage vacuum crystallization system from which the pure melamine is recovered in a continuous centrifuge. Stainless steel is used as construction material for nearly all parts of the equipment exposed to product streams.

Analysis and Toxicity. Melamine is determined gravimetrically by precipitation of the insoluble oxalate from a hot aqueous solution (63).

Extensive toxicity investigations performed with melamine in experimental animals suggest that the compound may have a low order of biological activity. The acute oral LD₅₀ for mice was found to be 4.55 g/kg and approximately the same for rats. Chronic feeding tests have been carried out on rats over a two-year period at a dietary level of 1000 ppm and on dogs for one year at a level of 30,000 ppm. Throughout the study, the general health of the test animals did not seem significantly different from that of the controls. Nevertheless, after 60–90 days the dogs showed melamine crystalluria, which persisted throughout the remainder of the observation. However, gross and microscopic examination of the tissues revealed no abnormality attributable to the feeding of melamine.

Melamine in a skin test on rabbits produced neither local irritation nor systemic toxicity. As a 10% solution in methylcellulose, it caused no irritation in the eyes of rabbits. Human subjects were given patch tests with melamine. No evidence of either primary irritation or sensitization was found. Such results suggest that melamine crystal may be handled in ordinary industrial use without special hygienic precautions.

Uses. Most of the melamine produced is used in the form of melamine–formaldehyde resins (see AMINO RESINS AND PLASTICS). Other applications (63) include the use of melamine pyrophosphate [15541-60-3] in fire retardant textile finishes, chlorinated melamine as a bactericide, and melamine as a tarnish inhibitor in detergent compositions, in papermaking, and manufacture of adhesives.

BIBLIOGRAPHY

"Cyanamides" in *ECT* 1st ed., Vol. 4, pp. 663–675, by V. Alexanderson, American Cyanamid Co., and L. M. Sherman, Chemical Construction Corp.; in *ECT* 2nd ed., Vol. 6, pp. 553–573, by J. R. McAdam, Cyanamid of Canada Ltd., and F. C. Schaefer, American Cyanamid Co.; in *ECT* 3rd ed., Vol. 7, pp. 291–306, by D. R. May, Cyanamid Canada Inc.

1. B. E. Turner and co-workers, *Astrophys. J.* **201**, 3 (1975).
2. D. Costa and C. Bolis-Cannella, *Ann. Chim. (Rome)*, **43** (1953).
3. W. H. Fletcher, *J. Chem. Phys.* **39**, 2478 (1963).
4. *Cyanamide*, technical bulletin, American Cyanamid Co., Wayne, N.J., 1966.
5. K. J. Odo, *J. Chem. Soc. (Japan) Pure Chem. Sect.* **74** (1953).
6. F. Baum, *Biochem. Z.* **26**, 325 (1910).
7. N. Kameyama, *Trans. Am-Electrochem. Soc.* **40**, 131 (1921); S. Kawai, *Sci. Pap. Inst. Phys. Chem. Res.* **16**, 24, 306 (1931).
8. G. Grude and G. Motz, *Z. Phys. Chem.* **118**, 145 (1925); U.S. Pat. **2,443,504** (June 15, 1948), I. Hechenbleikner (to American Cyanamid Co.).
9. E. Mulder and J. A. Roorda-Smit, *Ber.* **7**, 1634 (1874).
10. L. A. Pinck and H. C. Hetherington, *Ind. Eng. Chem.* **18**, 629 (1926).
11. M. Kurabayashi and K. Yanagiya, *Rep. Govt. Chem. Eng. Res. Inst. (Tokyo)* **48**, 133 (1953).
12. H. Pincas, *Ind. Chim.* **21**, 413 (1934).
13. S. A. Miller and B. Bann, *J. Appl. Chem. (London)* **6**, 89 (1956).
14. *Methods of Analysis*, Association of Official Agricultural Chemists, Washington, D.C., 1955; **Can. Pat.** **956,081** (Oct. 15, 1974), F. C. Schaefer (to American Cyanamid Co.).
15. R. H. McKee, *Am. Chem. J.* **36**, 208 (1906); U.S. Pat. **2,476,452** (July 19, 1949), J. H. Paden and A. F. MacLean (to American Cyanamid Co.).
16. U.S. Pat. **2,727,922** (Dec. 20, 1955), H. Z. Lecher and C. L. Kosloski (to American Cyanamid Co.).
17. S. Basterfield and co-workers, *Can. J. Res.* **1**, 261 (1929); F. Kurzer and A. Lawson, *Org. Syn.* **34**, 67 (1954).
18. M. L. Kasten and W. G. McBurney, *Ind. Eng. Chem.* **43**, 1020 (1951).
19. B. A. Peacock, American Cyanamid co., Whitbank, SA, July 1991.
20. F. Riten and J. Krause, in K. Winnaker and L. Kuchler, eds., *Chemische Technologie*, Vol. 2, 2nd ed., Carl Hanser Verlag, Munich, Germany, 1959, p. 269.
21. U.S. Pat. **2,838,379** (Sept. 15, 1958), F. Kaess and co-workers (to Suddeutsche Kalkstickstoff Werke).
22. U.S. Pat. **3,300,281** (Jan. 24, 1967), D. R. May (to American Cyanamid Co.).
23. U.S. Pat. **3,295,926** (Jan. 3, 1967), R. H. Janes, G. J. Novak, and R. C. Rawlings (to American Cyanamid Co.).
24. *Cyanamide by Cyanamid*, American Cyanamid Co., Wayne, N.J., 1961; O. I. Polyanshikov and co-workers, *Vopr. Khim. Khim. Tekhnol.* **39**, 136 (1975).
25. A. A. Pimenova and co-workers, *Tr. Tashk. P'olitekh. Inst.* **107**, 49 (1973).
26. V. G. Golov and co-workers, *Zh. Khim. Abstr.* 23V1s (1975).
27. L. A. Pinck, *Ind. Eng. Chem.* **17**, 459 (1925).
28. *Determination of Nitrogen by FP-228 Organic Nitrogen Terminator*, LECO Corp., St. Joseph, Mich., 1985.
29. G. R. Dixon and F. Wilson, *Tests Agro Chem. Culti* **4**, 50–1, (1983).
30. A. M. A. Al-Wakeel, and A. M. A. Ismail, *Inst. Tiermed Tierhyg.* **94**(7), 131–133 (1981).
31. R. Heitefuss and R. R. Raeschke, *Herbizide*, 57–62 (1979).
32. U.S. Pat. **4,487,625** (Dec. 11, 1984), G. Rieder (to SKW, Germany).
33. K. Arikawa and K. Inanaga, *Folia Psychiat. Neurol. Jpn.* **27**(1), 9 (1973).
34. **Ger. Pats.** **1,771,827** (May 24, 1973), P. Birk and K. Wohlgemuth (to Goerig & Co.); **2,136,450** (Feb. 1, 1973), K. Deutzmann (to W. Banck and O. Vorbach).
35. **Ger. Pats.** **2,252,795** (May 22, 1974), W. Meichsner (to Suddeutsche Kalkstickstoff-Werke A.G.); **2,252,796** (May 22, 1973), A. Freissmuth (to Suddeutsche Kalkstickstoff-Werke A.G.).

36. **Ger. Pat. 2,031,720** (Jan. 5, 1972), H. Prietzel (to Suddeutsche Kalkstickstoff-Werke A.G.).

37. **Belg. Pat. BE 897848 A1** (Mar. 28, 1984), (to Produits Chimiques Ugine Kuhlmann).

38. **U.S. Pat. 4,000,302** (Dec. 28, 1976), J. W. Black (to Smith Kline).

39. **Ger. Pat. 2,460,330** (July 1, 1976), K. Lohmer (to Agfa).

40. N. Kameyama, *J. Chem. Ind. (Japan)* **24**, 1263 (1921).

41. *AERO Dicyandiamide*, American Cyanamid Co., Wayne, N.J., 1964.

42. *SKW-Produktstudie Dicyandiamid*, Suddeutsche Kalkstickstoff-Werke A.G., Trostberg, Germany, 1973, 60 pp.

43. T. C. Ponder, *Hydrocarbon Process*, **48**(11), 200 (1969).

44. P. Ostrogovich, *Atti Accad. Lincei* **20**(1), 249 (1911); **U.S. Pat. 2,302,162** (Nov. 17, 1942), W. Zerwick and W. Brunner (vested in the Alien Property Custodian).

45. F. H. S. Curd and F. L. Rose, *Chem. Ind.* **75** (1946); **Brit. Pat. 581,346** (1946), F. H. S. Curd and F. L. Rose.

46. J. H. Paden, K. C. Martin, and R. C. Swain, *Ind. Eng. Chem.* **39**, 952 (1947).

47. J. Haag, *Ann. Chem.* **122**, 22 (1862).

48. **Brit. Pat. 434,961** (Sept. 12, 1935) (to E. I. du Pont de Nemours & Co., Inc.).

49. C. Hasegawa, *J. Soc. Chem. Ind. (Japan)* **45**, 416 (1942).

50. C. D. Garby, *Ind. Eng. Chem.* **17**, 266 (1925).

51. R. N. Harger, *Ind. Eng. Chem.* **12**, 1107 (1920); E. Johnson, *Ind. Chem.* **13**, 533 (1921).

52. J. W. Lyons, *The Chemistry and Use of Fire Retardants*, John Wiley & Sons, Inc., New York, 1970, p. 136.

53. **U.S. Pat. 3,159,503** (Dec. 1, 1964), I. S. Goldstein and W. A. Dreher (to Koppers Co., Inc.).

54. **U.S. Pats. 2,280,995** (Apr. 28, 1942), R. B. Booth (to American Cyanamid Co.); **2,581,111** (Jan. 1, 1952), C. G. Landes and J. Studeny (to American Cyanamid Co.).

55. **U.S. Pat. 2,523,626** (Sept. 26, 1950), G. Jones, W. Juda, and S. Soll (to Albi Manufacturing Co., Inc.).

56. **Jpn. Pat. 59/136368 A2** (Aug. 4, 1984) (to Toshiba Chemical Products Co. Ltd.).

57. **U.S. Pat. 4,537,805** (Aug. 27, 1985), Shiow C Lin (to W. R. Grace and Co.).

58. **USSR Pat. SU 1000511 A1** (Feb. 28, 1983), M. V. Frolov, V. N. Chemyslev, V. I. Volovich, Yu Toporov (to Central Scientific Research Institute of Paper.

59. G. DiModica and A. Tortosa, *Rev. Tech. Ind. Cuir* **74**(11), 3 63–7 (1982).

60. **USSR Pat. 975,763 A1** (Nov. 23, 1982), A. N. Gromov, M. G. Kodzhoglyan, Z. G. Akimova, and A. I. Petrov, (to USSR).

61. **Jpn. Pat. 57/147513 A2** (Sept. 11, 1982) (to Sumitomo Chemical Co. Ltd.).

62. **Jpn. Pat. 60/69128 A2** (Apr. 19, 1985) (to Nitto Electric Industries Co. Ltd.).

63. *AERO Melamine*, American Cyanamid Co., Wayne, N.J., 1968.

64. B. Bann and S. A. Miller, *Chem. Res.* **58**, 131 (1958).

Baldev K. Patel
Cyanamid Canada Inc.

CYANIDES

HYDROGEN CYANIDE

Hydrogen cyanide [74-90-8] (hydrocyanic acid, prussic acid, formonitrile), HCN, is a colorless, poisonous, low viscosity liquid having an odor characteristic of bitter almonds. It was prepared in dilute solution by Scheele in 1782 and as the pure compound by Gay-Lussac in 1815. The compound has been known and used as a poison for decades (see POISONS, COMMERCIAL). In England, hydrogen cyanide was sold in small crates of glass bottles for the fumigation of old furniture being rehabilitated for sale as antiques. In the United States, hydrogen cyanide was used for fumigation in the orange groves in California. It was transported to the groves by horse-drawn carts in milk cans. Although until the 1960s hydrogen cyanide was used as a fumigant in grain silos, this use has virtually disappeared. Hydrogen cyanide is used, however, in the manufacture of many important chemicals.

It is theorized that hydrogen cyanide played the key role in the origin of plant and animal life on earth via formation of amino acids (1,2). Hydrogen cyanide is present in the normal human being's blood. People who smoke or who consume vegetables having relatively high cyanide content have slightly higher blood concentrations.

Some hydrogen cyanide is formed whenever hydrocarbons (qv) are burned in an environment that is deficient in air. Small concentrations are also found in the stratosphere and atmosphere. It is not clear whether most of this hydrogen cyanide comes from biological sources or from high temperature, low oxygen processes such as coke production, but no accumulation has been shown (3).

In early times hydrogen cyanide was manufactured from beet sugar residues and recovered from coke oven gas. These methods were replaced by the Castner process in which coke and ammonia were combined with liquid sodium to form sodium cyanide. If hydrogen cyanide was desired, the sodium cyanide was contacted with an acid, usually sulfuric acid, to liberate hydrogen cyanide gas, which was condensed for use. This process has since been supplanted by large-scale plants, using catalytic synthesis from ammonia and hydrocarbons.

Hydrogen cyanide is a basic chemical building block for such chemical products as adiponitrile to produce nylon, methyl methacrylate to produce clear acrylic plastics (see ACRYLIC ESTER POLYMERS), sodium cyanide for recovery of gold (see GOLD AND GOLD COMPOUNDS), triazines for agricultural herbicides (qv), methionine for animal food supplement (see FEEDS AND FEED ADDITIVES), chelating agents (qv) for water treatment, and many more.

Properties

The physical properties of hydrogen cyanide are listed in Table 1.

Table 1. Physical Properties of Hydrogen Cyanide

Property	Value
molecular weight	27.03
melting point, °C	13.24
triple point, °C	13.32
boiling point, °C	25.70
density, g mL	
0°C	0.7150
10°C	0.7017
20°C	0.6884
specific gravity of aqueous solutions ^a	
10.04° º HCN	0.9838
20.29° º HCN	0.9578
60.23° º HCN	0.829
vapor pressure, kPa ^b	
29.5°C	6.697
0°C	35.24
27.2°C	107.6
vapor specific gravity, at 31°C ^c	0.947
surface tension at 20°C, mN m (dyn m)	19.68
liquid viscosity at 20.2°C, mPa·s (cP)	0.2014
specific heat, J mol ^d	
33.1°C, liquid	58.36
16.9°C, liquid	70.88
27°C, gas	36.03
heat of fusion at -14°C, kJ mol ^d	7.1 × 10 ³
heat of formation, ΔH _f , kJ mol ^d	
gas at 25°C	130.5
liquid at 25°C	105.4
heat of combustion, net, kJ mol ^d	642
critical temperature, °C	183.5
critical density, g mL	0.195
critical pressure, MPa ^e	5.4
dielectric constant	
0°C	158.1
20°C	114.9
dipole moment, gas, at 3–15°C, Cm ^f	7.0 × 10 ⁻³⁰

ionization constant, K, at 25°C	7.2 × 10 ⁻¹⁰
conductivity, S cm	3.3 × 10 ⁻⁶
heat of vaporization, kJ mol ^d	25.2
heat of polymerization, kJ mol ^d	42.7
entropy, gas at 27°C, 100 kPa ^b , J (mol°C) ^d	202.0
flash point, closed cup, °C	17.8
explosive limits in air at 100 kPa ^b and 20°C, vol %	6–41
autoignition temperature, °C	538
index of refraction, liquid at 10°C	1.2675

^a Measured at 18°C, compared to water at 18°C.

^b To convert kPa to mm Hg, multiply by 7.5.

^c Air 1.

^d To convert J to cal, divide by 4.184.

^e To convert MPa to atm, divide by 0.101.

^f To convert Cm to debye, divide by 3.336 × 10⁻³⁰.

Chemical Properties. Hydrogen cyanide is a weak acid; its ionization constant is of the same magnitude as that of the natural amino acids (qv). Its structure is that of a linear, triply bonded molecule, HC≡N.

Hydrogen cyanide, as the nitrile of formic acid [64-18-6], CH₂O₂, undergoes many of the typical nitrile reactions. For example, it can be hydrolyzed to formic acid by aqueous sulfuric acid (4); it can be hydrogenated to methylamine [74-89-5], CH₅N (5); and it can be converted to phenylformamidine [618-39-3], C₇H₈N₂, using aniline and hydrogen chloride (6).

Hydrogen cyanide can be oxidized by air at 300–650°C over silver (7) or gold (8) catalyst to give yields of up to 64% cyanic acid [420-05-3], HOCN, and 26% cyanogen [460-19-5], (CN)₂. Reaction with chlorine in the liquid phase gives cyanogen chloride [506-77-4], CClN (9,10), which is the basic route to triazines of which melamine [108-78-1], C₃H₃N₃, is an important derivative (see CYANAMIDES; UREA). Bromine reacts similarly, but the reaction with iodine is incomplete. Chlorination in the vapor phase can be controlled to give cyanogen (11), cyanogen chloride, or cyanuric chloride [108-77-0], C₃Cl₃N₃ (12) (see CYANURIC AND ISOCYANURIC ACIDS).

Hydrogen cyanide adds to an olefinic double bond most readily when an adjacent activating group is present in the molecule, eg, carbonyl or cyano groups. In these cases, a Michael addition proceeds readily under basic catalysis, as with acrylonitrile (qv) to yield succinonitrile [110-61-2], C₄H₄N₂, in high yield (13). Formation of acrylonitrile by addition across the acetylenic bond can be accomplished under catalytic conditions (see ACETYLENE DERIVED CHEMICALS).

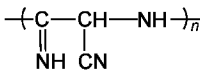
Hydrogen cyanide adds across the carbonyl group of aldehydes (qv) and ketones (qv) and opens the oxirane ring of epoxides, both under mildly basic conditions. Several of these cyanohydrins (qv) are commercially important. Acetone cyanohydrin [75-86-5] is an intermediate in the manufacture of methyl methacrylate for acrylic plastics (see METHACRYLIC ACID AND DERIVATIVES). Lactic acid [598-82-3], C₃H₅O₃, is made by the hydrolysis of lactonitrile [78-97-7], C₃H₅NO, which is formed from the reaction of acetaldehyde (qv) and hydrogen cyanide. Hydrogen cyanide reacts with phenols in the presence of hydrogen chloride and aluminum chloride to give aromatic aldehydes (the Gattermann synthesis). Hydrogen cyanide is a starting chemical for alanine [302-72-7], phenylalanine [63-91-2], valine [72-18-4], and aminobutyric acid.

Hydrogen cyanide reacts with formaldehyde and aniline to form *N*-phenylglycinonitrile [3009-97-0], and with formaldehyde alone to form glycolonitrile [107-16-4] (14). Hydrogen cyanide reacts with NaOH, KOH, and Ca(OH)₂ to form the corresponding cyanides. Amines can be derived from olefins and hydrogen cyanide via the Ritter reaction (15). Nitrilotriacetoneitrile [628-87-5], C₄H₅N₃, is produced by the reaction of a salt of ammonia and a nonoxidizing acid with formaldehyde and hydrogen cyanide (16). Cyanogen can be formed by oxidative cleavage of hydrogen cyanide over a silver catalyst (17). Melamine can be prepared from hydrogen cyanide and NH₃ by way of cyanamide (18). Iminodiacetonitrile [628-87-5] can be prepared from hexamethylenetetramine, formaldehyde, and hydrogen cyanide (19). Oxamide [471-46-5] is formed from hydrogen cyanide and hydrogen peroxide (20). Malononitrile [109-77-3] can be produced from cyanogen chloride, acetonitrile, and hydrogen cyanide at 700°C (21). Diiminosuccinonitrile is formed from hydrogen cyanide, chlorine, and trimethylamine (22).

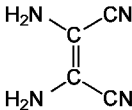
Cyanohydrins (qv) are formed by the reaction of glucose and similar compounds with hydrogen cyanide. The corresponding aminonitrile from methyl isobutyl ketone can be formed with ammonia and hydrogen cyanide.

Dimethylformamide [68-12-2] can be produced from the reaction of hydrogen cyanide and methanol. Adenine [73-24-5] can be prepared from hydrogen cyanide in liquid ammonia. Thioformamide [115-08-2] can be produced from hydrogen cyanide and hydrogen sulfide.

Under certain conditions hydrogen cyanide can polymerize to black solid compounds, eg, hydrogen cyanide homopolymer [26746-21-4] (1) and hydrogen cyanide tetramer [27027-02-2], C₄H₄N₄ (2). There is usually an incubation period before rapid onset of polymer formation. Temperature has an inverse logarithmic effect on the incubation time. Acid stabilizers such as sulfuric and phosphoric acids prevent polymerization. The presence of water reduces the incubation period.



(1)



(2)

Although hydrogen cyanide is a weak acid and is normally not corrosive, it has a corrosive effect under two special conditions: (1) water solutions of hydrogen cyanide cause transcrystalline stress cracking of carbon steels under stress even at room temperature and in dilute solution and (2) water solutions of hydrogen cyanide containing sulfuric acid as a stabilizer severely corrode steel (qv) above 40°C and stainless steels above 80°C.

Manufacture and Processing

Hydrogen cyanide has been manufactured from sodium cyanide and mineral acid, and from formamide by catalytic dehydration. As of this writing,

primarily because of high raw material costs, only one manufacturer uses the formamide route.

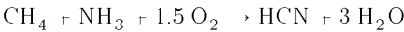
Two synthesis processes account for most of the hydrogen cyanide produced. The dominant commercial process for direct production of hydrogen cyanide is based on classic technology (23–32) involving the reaction of ammonia, methane (natural gas), and air over a platinum catalyst; it is called the Andrussov process. The second process involves the reaction of ammonia and methane and is called the Blaschke-Methan-Ammoniak (BMA) process (30,33–35); it was developed by Degussa in Germany. Hydrogen cyanide is also obtained as a by-product in the manufacture of acrylonitrile (qv) by the ammoxidation of propylene (Sohio process).

The Shawinigan process uses a unique reactor system (36,37). The heart of the process is the fluohmic furnace, a fluidized bed of carbon heated to 1350–1650°C by passing an electric current between carbon electrodes immersed in the bed. Feed gas is ammonia and a hydrocarbon, preferably propane. High yield and high concentration of hydrogen cyanide in the off gas are achieved. This process is presently practiced in Spain, Australia, and South Africa.

A process has been developed to synthesize hydrogen cyanide from coal and ammonia. This process has not yet been commercialized (38). Limited quantities of hydrogen cyanide are also recovered from coke oven gases. Hydrogen cyanide is produced when hydrogen, nitrogen, and carbon-containing compounds are brought together at high temperatures with or without a catalyst.

Examples of the variety of systems that yield hydrogen cyanide but that have not yet shown economic significance include: at 1100–1300°C acetonitrile and ammonia produce a gas containing 47.3 mol % HCN, 1.1 mol % NH₃, 0.8 mol % NO, 0.4 mol % CH₄, 0.04 mol % CH₃CN, and 49.9 mol % H₂ (39); hydrogen cyanide is produced from the reduction of nitric oxide over precious metal and perovskite catalysts at 400–800°C (40); and methanol and ammonia react in the absence of O₂ and in the presence of catalysts at 600–950°C to produce hydrogen cyanide (41).

The stoichiometry of the Andrussov process may be represented by



$$\Delta H = +481.9 \text{ kJ mol}^{-1} (115.2 \text{ kcal mol}^{-1})$$

The oxidation of the hydrogen is not complete so that the converter off-gas contains hydrogen. The overall reaction is carried out adiabatically. This is accomplished by the addition of air (O₂). The air oxidizes a portion of the methane, making the overall reaction exothermic, even though the reaction of methane with ammonia to form hydrogen cyanide is quite endothermic.

The catalyst temperature is about 1100°C. Precious metal catalysts (90% Pt 10% Rh in gauze form) are normally used in the commercial processes. The converters are similar to the ammonia oxidation converters used in the production of nitric acid (qv) although the latter operate at somewhat lower temperatures. The feed gases to the converter are thoroughly premixed. The optimum operating mixture of feed gas is above the upper flammability limit and caution must be exercised to keep the mixture from entering the explosive range.

The reactions taking place in the Andrussov process are more complex than that shown (42). Most of the heat required for hydrogen cyanide formation is supplied by combustion of methane. The endotherm of the methane–ammonia reaction is 251 kJ mol (60 kcal mol) of hydrogen cyanide formed. Preventing the decomposition of ammonia and hydrogen cyanide is a critical aspect of successful operation. The equilibrium concentration of hydrogen cyanide in the converter gases at 1100°C is 0%. Therefore, to prevent hydrogen cyanide decomposition, the converter gases must be quickly quenched to <400°C. This is done in a steam-generating waste-heat boiler located directly below the catalyst gauze.

Modern hydrogen cyanide plants are energy cost sensitive, and recovery of heat is important. The waste-heat boiler directly below the catalyst bed accomplishes two things: (1) the gases are quenched to <400°C to minimize hydrogen cyanide decomposition and (2) steam is generated to recover energy. Approximately 5 kg of steam is generated per kilogram of hydrogen cyanide. Conversion, yields, and productivity of the synthesis unit are influenced by the extent of feed gas preheat, purity of the feeds, reactor geometry, feed gas composition, contact time, catalyst composition and purity, converter gas pressure, quench time, and materials of construction.

In one patent (31), a filtered, heated mixture of air, methane, and ammonia in a volume ratio of 5:1:1 was passed over a 90% platinum–10% rhodium gauze catalyst at 200 kPa (2 atm). The unreacted ammonia was absorbed from the off-gas in a phosphate solution that was subsequently stripped and refined to 90% ammonia–10% water and recycled to the converter. The yield of hydrogen cyanide from ammonia was about 80%. On the basis of these data, the converter off-gas mol % composition can be estimated: nitrogen, 49.9%; water, 21.7%; hydrogen, 13.5%; hydrogen cyanide, 8.1%; carbon monoxide, 3.7%; carbon dioxide, 0.2%; methane, 0.6%; and ammonia, 2.3%.

There are two processes for recovering the hydrogen cyanide from the converter off-gases. One, shown in Figure 1, recovers unreacted ammonia for recycle to the converter, and the other, shown in Figure 2, scrubs the ammonia out of the gas using sulfuric acid, and produces ammonium sulfate [7783-20-2], (NH₄)₂SO₄, as a by-product. This ammonium sulfate can be a disposal problem, but the recycle system is rather capital and energy intensive. Ammonia must be removed from the off-gas before hydrogen cyanide can be recovered as the ammonia promotes polymerization of the hydrogen cyanide. In processes where ammonia is recovered as ammonium sulfate, it can be processed to dry crystalline ammonium sulfate.

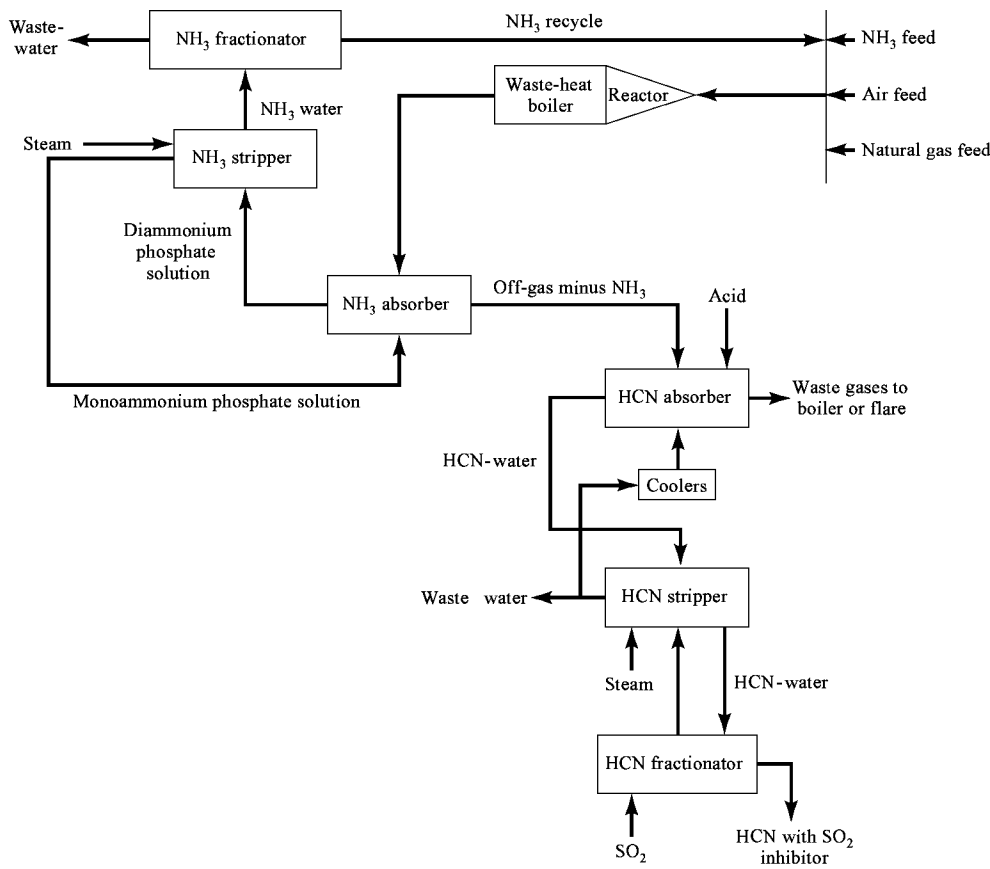


Fig. 1. Andrussow process plant with ammonia recycle.

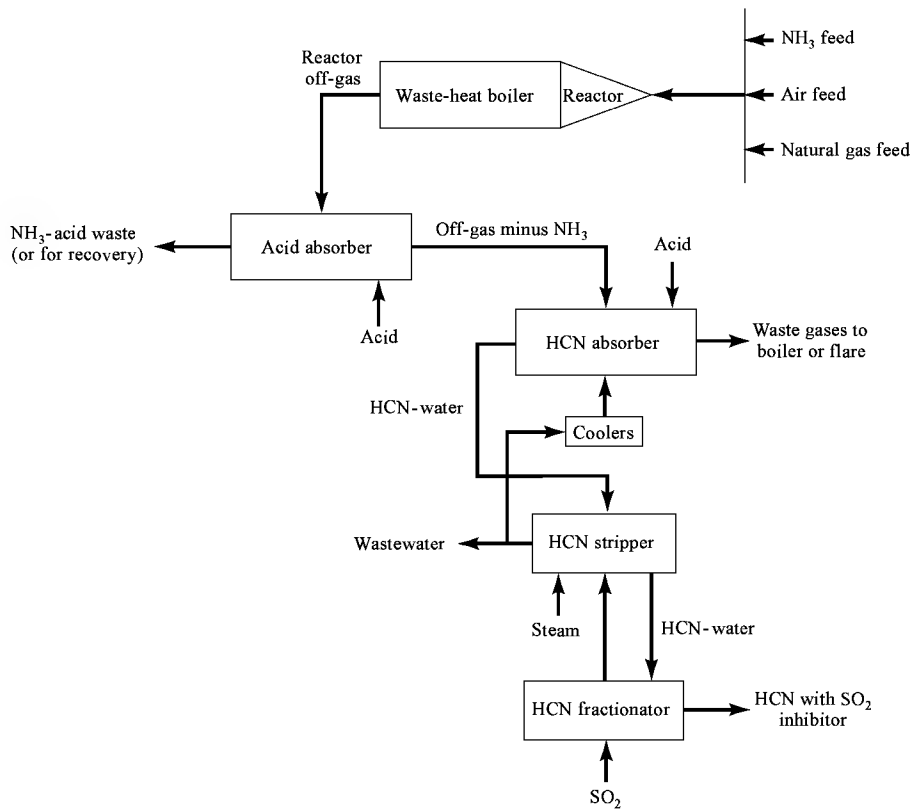


Fig. 2. Andrussow process without ammonia recycle.

The approach that recycles ammonia to the converter involves passing the off-gas from the converters through a solution of monoammonium phosphate at 60–80°C where the ammonia reacts to form diammonium phosphate. The diammonium phosphate solution is stripped back to monoammonium phosphate with steam and the ammonia–water overheads are fractionated under pressure to provide ammonia gas suitable for recycle to the converter (43). The ammonia-free off-gas from the monoammonium phosphate absorber passes to a second absorber for absorption of hydrogen cyanide in cold water. The dilute solution is stripped and fractionated by conventional means to 99.5% or higher purity. The absorber and downstream equipment are acidified with a trace of acid to avoid polymerization. A trace of sulfur dioxide is also added to the product hydrogen cyanide, which acts as stabilizer in the areas where hydrogen cyanide vapor might condense on tank or vessel ceilings. Sulfuric acid is added for supplemental stabilization. To avoid corrosion problems, all recovery equipment is constructed of austenitic stainless steel. Trace quantities of the following impurities are present in the product: cyanogen, acrylonitrile, acetonitrile, and propionitrile.

The waste gas remaining after removal of ammonia and recovery of hydrogen cyanide contains enough hydrogen and carbon monoxide that it is flammable and has enough heat value to make it a valuable fuel. It is usually used to displace other fuel in boilers.

Air pollution from this process is a minimal problem because all waste gases are burned for fuel value or flared. The products of combustion of the

waste gases are primarily water from the hydrogen and some carbon dioxide from the carbon monoxide. Aqueous discharge is essentially the converter water of reaction, which is contaminated by traces of hydrogen cyanide and other nitriles and the ammonium sulfate. The most common treatment for this aqueous stream is alkaline chlorination, which converts cyanide to the essentially nontoxic cyanate.

The Andrussow process is by far the dominant hydrogen cyanide manufacturing process in use, even though the converter off-gas stream is dilute in hydrogen cyanide. This forces the absorbers to be large so that they can handle the large volumes of gas. Preheating Andrussow feed gases is claimed to result in an increased hydrogen cyanide concentration by supplying part of the required reaction heat without *in situ* combustion (32). Advantages of the Andrussow process are low converter investment, low maintenance costs, and high natural gas yields when the waste gas is used as fuel.

In the BMA process, methane (natural gas) and ammonia are reacted without air being present (44). The reaction is carried out in tubes that are heated externally to supply the endothermic heat of reaction very similar to a reformer. Yield from ammonia and methane is above 90%. The off-gas from the converter contains more than 20 mol % hydrogen cyanide, about 70 mol % hydrogen, 3 mol % ammonia, 1 mol % methane, and about 1 mol % nitrogen from ammonia decomposition.

A typical converter is made up of multiple furnaces, each of which contains 8 to 10 reactors. Each reactor is made up of 10 to 30 sintered alumina tubes lined with platinum. The furnaces are direct fired with natural gas to 1200–1300°C. A typical furnace can produce about 125 t per month of hydrogen cyanide. Catalyst life is approximately 10,000 h.

After removal of the unreacted ammonia and recovery of hydrogen cyanide, the waste gas is essentially all hydrogen suitable for other chemical use. The advantages of the BMA process are the high ammonia and natural gas yields and the useful hydrogen waste gas, but the high investment and maintenance for the converter is a decided disadvantage.

A significant source of hydrogen cyanide (about 30% worldwide) is as by-product from the Sohio process acrylonitrile plants. Approximately 15 kg of HCN is produced per 100 kg acrylonitrile (qv). Catalysts have been developed that improve efficiency by decreasing the amount of hydrogen cyanide produced per kilogram of acrylonitrile. This, along with the declining demand for acrylonitrile, has resulted in an ever increasing need for on-purpose hydrogen cyanide.

The fluohmic process is a third process for manufacturing hydrogen cyanide, which is being applied in Spain and Australia. This process involves the reaction of ammonia with a hydrocarbon, usually propane or butane, in a fluidized bed of coke particles. The endothermic heat of reaction is supplied electrically through electrodes immersed in the fluid bed. Yields from propane and ammonia are reportedly above 85% and the waste gas is essentially hydrogen, but the costs for electricity are high. Thus this process is applicable only when there is an inexpensive source of power.

Recovery of hydrogen cyanide from coke-oven gases has been dormant in the early 1990s, but new methods involving environmental control of off-gas pollutants may be leading the way for a modest return to the recovery of cyanide from coke-oven gases (see COAL CONVERSION PROCESS, CARBONIZATION).

Economic Aspects

Annual U.S. hydrogen cyanide capacity in 1991 was 734,000 t as shown in Table 2. U.S. production for 1983–1989 has been estimated as shown in Table 3. Output of hydrogen cyanide in the United States rose to 600,000 t yr in 1992. Worldwide annual production and capacity of hydrogen cyanide in 1992 were estimated to be 950,000 and 1,320,000 t, respectively.

Table 2. U.S. Producers of Hydrogen Cyanide and Capacities

Producer	Capacity, 10 ³ t yr
American Cyanamid, Fortier, La. ^a	18.0
BP Chemicals ^a	50.0
Ciba Geigy Corp.	41.0
Degussa Corp.	24.0
DOW Chemical U.S.A.	9.0
E. I. du Pont de Nemours & Co., Inc.	410.0
Monsanto Co. ^a	30.0
Rohm and Haas Co.	90.0
Stedling Chemical ^a	40.0
Cyanco Co.	7.0
FMC Corp.	15.0
<i>Total</i>	<i>734.0</i>

^a HCN is a by-product.

Table 3. U.S. Production of Hydrogen Cyanide for 1983 to 1989

Year	Production, 10 ³ t yr
1983	330
1984	365
1985	365
1986	430
1987	470
1988	500
1989	490

Specifications and Analysis

Hydrogen cyanide is classified by the U.S. Department of Transportation as a Class A Flammable Poison and is subject to rigid packaging, labeling, and shipping regulations. Hydrogen cyanide can be purchased in cylinders with capacities ranging from 300 mL up to 75 kg. Purchases may be made in tank car sizes of 24 and 46 t at \$1.30 kg. Although there have been no accidents since the inception of bulk transportation (since 1950), the continuing trend toward reducing risk from hazardous chemicals threatens the transportation of hydrogen cyanide.

Specifications are 99.5 wt % hydrogen cyanide (min), 0.5 wt % water (max), 0.06–0.10% acidity, and color not darker than APHA 20. A combination of H₂SO₄ (or H₃PO₄) and SO₂ acts as a stabilizer to prevent polymerization; H₂SO₄ stabilizes the liquid phase and SO₂ stabilizes the vapor phase.

Assay of hydrogen cyanide can be done by specific gravity or silver nitrate titration. Sulfur dioxide in hydrogen cyanide can be determined by infrared analysis or by reaction of excess standard iodine solution and titration, using standard sodium thiosulfate or by measurement of total acidity by

evaporation to dryness and remeasurement of acidity. Total acidity, which includes SO₂ plus H₂SO₄ or H₃PO₄, can be obtained by titration using 0.1 N NaOH using methyl red–methylene blue mixed indicator. Water can be determined by Kad Fischer titration.

Health and Safety Factors

The cyanides are true noncumulative protoplasmic poisons, ie, they can be detoxified readily. Cyanide combines with those enzymes at the blood tissue interfaces that regulate oxygen transfer to the cellular tissues. Unless the cyanide is removed, death results through insufficient oxygen in the cells. The warning signs of cyanide poisoning include dizziness, numbness, headache, rapid pulse, nausea, reddened skin, and bloodshot eyes. More prolonged exposure can cause vomiting and labored breathing followed by unconsciousness; cessation of breathing; rapid, weak heart beat; and death. Severe exposure by inhalation can cause immediate unconsciousness; this rapid knockdown power without an irritating odor makes hydrogen cyanide more dangerous than materials of comparable toxicity, eg, H₂S. Hydrogen cyanide can enter the body by inhalation, oral ingestion, or skin absorption.

Inhalation. The threshold limit value of HCN is 4.7 ppm. This is defined as the maximum average safe exposure limit for a 15-min period by the Occupational Safety and Health Administration. Exposure to 20 ppm of HCN in air causes slight warning symptoms after several hours; 50 ppm causes disturbances within an hour; 100 ppm is dangerous for exposures of 30 to 60 min; and 300 ppm can be rapidly fatal unless prompt, effective first aid is administered. There is always a small concentration of cyanide (0.02 to 0.04 mg L) in the blood, and the body has a mechanism for continuous removal of small amounts, such as from smoking, by converting it to thiocyanate, which is discharged in the urine.

Skin Absorption. Normal skin absorbs HCN slowly. However, 2° ° HCN in air may cause poisoning in 3 min, 1° ° is dangerous in 10 min, and 0.05° ° may produce symptoms after 30 min, even though a gas mask or air mask is worn. Some areas of the body, such as the feet and mucous membranes, are more absorptive than others. Cuts and abrasions absorb cyanide rapidly, and 50 mg of HCN absorbed through the skin can be fatal.

Ingestion. Ingestion, unless prompt first aid or medical treatment is given, is rapidly fatal; 1 mg of cyanide per kilogram of body weight can be fatal. Immediate and repeated administration of emetics and regurgitation (if the victim is conscious), followed or accompanied by the first aid and medical treatments described below should be carried out. If the victim is unconscious, stomach lavage should be performed by a physician or trained personnel.

First Aid and Medical Treatment. Before using or handling cyanides, intensive training in the proper first-aid protocol is essential. In case of an accident, action should be fast and efficient. With the protection of a gas mask remove or drag the victim to fresh air. Remove contaminated clothing and rinse contaminated body areas. Keep victim warm. If the victim is conscious and speaking, no treatment is necessary. If the victim is unconscious but breathing, break an ampul of amyl nitrite in a cloth and hold it under the victim's nose for 15 s. Repeat five or six times. Use a fresh ampul every 3 min. Continue until the victim regains consciousness. Amyl nitrite is a powerful cardiac stimulant and should not be used more than necessary. If the patient is not breathing, apply artificial respiration; this can best be done using an oxygen resuscitator. The amyl nitrite antidote should also be administered during resuscitation. Mouth-to-mouth resuscitation is the next-best method followed by the Holger-Mielsen arm-lift method.

Notify a physician immediately. A suggested procedure for physicians or nurses is intravenous administration of 0.3 g (10 mL of a 3° ° solution) of sodium nitrite at the rate of 2.5 mL min followed by 12.5 g (50 mL of a 25° ° solution) of sodium thiosulfate at the same rate. Watch the patient for 24 to 48 h, especially in cases of ingestion or skin absorption. If symptoms reappear, repeat the injections in half the original amounts. These solutions should be kept readily available. In some cases, first aid personnel have been trained to use the intravenous medication subject to government regulations.

Detection. Many people can detect hydrogen cyanide by odor or taste sensation at the 1 ppm concentration in air, most at 5 ppm, but HCN does not have an offensive odor and a few people cannot smell it even at toxic levels. Anyone planning to work with hydrogen cyanide should be checked with a sniff test employing a known safe concentration. This test should be given periodically. Several chemical detection and warning methods can be employed. The most reliable are modern, electronic monitors based on electrolytes that react with hydrogen cyanide.

Disposal. Small quantities of concentrated hydrogen cyanide can be burned in a hood in an open vessel. Large-scale burning in outdoor pans can be performed, but special safety precautions must be employed. A cyanide solution can be decontaminated by making the solution strongly basic (pH 12) with caustic and pouring it into ferrous sulfate solution. The resulting ferrocyanide is relatively nontoxic. Cyanide solution can be converted to less toxic cyanate by treatment with chlorine, sodium or calcium hypochlorite, or ozone at pH 9 to 11. A solution of 10° ° hypochlorite maximum should be used. The final solution should be checked for absence of free cyanide. The hypochlorite or Cl₂ + NaOH method is by far the most widely used commercially (45). However, other methods are oxidation to cyanate using hydrogen peroxide, ozone, permanganate, or chlorite; electrolysis to CO₂, NH₃, and cyanate; hydrolysis at elevated temperatures to NH₃ and salts of formic acid; air or steam stripping at low pH; biological decomposition to CO₂ and N₂; chromium treatment to give oxamides; nitrite or nitrate treatment to give carbonates and nitrogen; removal by ion exchange and recovery; lime–sulfur reaction to give sulfate, carbonate, and free sulfur; and gamma-ray treatment to give nontoxic constituents.

Environmental. The toxicity of cyanide in the aquatic environment or natural waters is a result of free cyanide, ie, as HCN and CN⁻. These forms, rather than complexed forms such as iron cyanides, determine the lethal toxicity to fish. Complexed cyanides may revert to free cyanide under uv radiation, but the rate is too slow to be a significant toxicity factor. Much work has been done to establish stream and effluent limits for cyanide to avoid harmful effects on aquatic life. Fish are extremely sensitive to cyanide, and the many tests indicate that a free cyanide stream concentration of 0.05 mg L is acceptable (46), but some species are sensitive to even lower concentrations.

Another important environmental issue is the fate of cyanide. Hydrogen cyanide, if spilled, evaporates quite readily. That which does not evaporate is soon decomposed or rendered nonhazardous by complexing with iron in the soil, biological oxidation, or polymerization.

General Safety Aspects. Whereas cyanides have become infamous as poisons in part because of the use in prison execution chambers, Nazi death camps, and in over-the-counter drug tampering cases, there is worldwide safe production of nearly one million t annually.

Laboratory work with hydrogen cyanide should be carried out only in a well-ventilated fume hood. Special safety equipment such as air masks, face masks, plastic aprons, and rubber gloves should be used. A chemical proof suit should be available for emergency. Where hydrogen cyanide is handled inside a building, suitable ventilation must be provided. The people involved should be thoroughly trained in first aid.

The most important rule when working with hydrogen cyanide is never to work alone. This applies especially to sampling and opening lines and equipment. A second person must be in view at all times about 9 to 10 m away, must be equipped to make a rescue, and must be trained in first aid for hydrogen cyanide exposure.

Besides toxicity, hydrogen cyanide presents other hazards. Hydrogen cyanide undergoes an exothermic polymerization at conditions of pH 5 to 11. This polymerization can become explosively violent, especially if confined. The reaction is between hydrogen cyanide and cyanide ions, so the presence of water and heat contribute to the onset of this polymerization. Stored hydrogen cyanide should contain less than 1 wt ° water, should be kept cool, and should be inhibited with sulfuric, phosphoric, or acetic acid. Manufacturers recommend a maximum of 90-day storage even for inhibited hydrogen cyanide. Cylinders should not be used as heated vaporizers as the elevated temperature accelerates the reaction of the stabilizing acid with the cylinder walls which can lead to an explosive polymerization. The presence of any contaminant or surface that depletes the acid stabilizer can cause polymerization, which is first indicated by a yellow-brown color in the hydrogen cyanide followed by development of heat. Prompt action to be taken includes reacidification and cooling. The resulting polymer is nontoxic except, of course, for any hydrogen cyanide that may be occluded in the polymer.

Explosively violent hydrolysis can occur if an excess of a strong acid (H₂SO₄, HNO₃, or HCl) is added to hydrogen cyanide. The reaction is fastest at or near stoichiometric ratios, eg, 1 to 2 moles H₂SO₄ per mole HCN, and can cause severe equipment damage if confined.

From a practical use standpoint there are some unique properties of hydrogen cyanide that have safety implications: (1) small concentrated liquid leaks form milk white icicles resulting from evaporative freezing, even in summer; (2) hydrogen cyanide has an unusually steep density change with temperature, eg, more than 10 times the change per degree as water; (3) CO₂ is soluble in hydrogen cyanide, which can cause hydrogen cyanide exposed to CO₂ during processing to foam and spout when warmed; and (4) this violent action also occurs when liquid hydrogen cyanide is depressurized from the superheated state. Because of its low boiling point, hydrogen cyanide can be a fire and explosion hazard.

Uses

Estimates of various uses for hydrogen cyanide in the United States are adiponitrile for nylon, 41° °; acetone cyanohydrin for acrylic plastics, 28° °; sodium

cyanide for gold recovery, 13° º; cyanuric chloride for pesticides and other agricultural products, 9° º; chelating agents such as EDTA, 4° º; and methionine [63-68-3] for animal feed, 2° º.

There has been a sizable shift in distribution of uses since the early 1980s. The use of hydrogen cyanide for manufacturing nylon via the adiponitrile route has expanded substantially and the use for producing sodium cyanide has also grown. Other uses have shown a more normal growth rate. Modest quantities of hydrogen cyanide go into a large number of relatively small uses, including manufacture of ferrocyanides (see IRON COMPOUNDS), acrylates, lactic acid, pharmaceuticals (qv), and specialty chemicals.

SODIUM CYANIDE

Sodium cyanide [143-33-9], NaCN, is a white cubic crystalline solid commonly called white cyanide. It was first prepared in 1834 by heating Prussian blue, a mixture of cyanogen compounds of iron, and sodium carbonate and extracting sodium cyanide from the cooled mixture using alcohol. Sodium cyanide remained a laboratory curiosity until 1887, when a process was patented for the extraction of gold and silver from ores by means of a dilute solution of cyanide (see METALLURGY, EXTRACTIVE). A mixture of sodium and potassium cyanides, produced by Erlenmeyer's improvement of the Rodgers process, was marketed in 1890.

The Beilby process started in 1891 and by 1899 accounted for half of the total European production of cyanide. In this process, a fused mixture of sodium and potassium carbonates reacts with ammonia in the presence of carbon. In 1900, the Castner process, in which molten sodium, ammonia, and charcoal react to give a high (98° º) grade sodium cyanide, superseded the Beilby process. Sodium cyanide became an article of commerce and soon replaced potassium cyanide in all except special uses.

The Castner process has been replaced by the neutralization or wet process in which liquid hydrogen cyanide and sodium hydroxide solution react and water is evaporated. The resulting crystals are briquetted or made into granular form. During the 1950s, essentially all sodium cyanide was used in electroplating (qv) and case hardening (see METAL SURFACE TREATMENTS). Upon the removal of control of the price of gold the use of cyanides for gold recovery expanded rapidly. Western gold production grew to 950 t in 1980 and continued to grow to 1470 t in 1991 (47). As of this writing, case hardening is a minor use. The principal applications of sodium cyanide are gold and silver extraction, electroplating, synthesis of iron blues, and synthesis of a large number of chemicals.

Properties

Physical Properties. The physical properties of sodium cyanide are listed in Table 4. The solid phase in contact with a saturated aqueous solution at temperatures above 34.7°C is the anhydrous salt; below 34.7°C, the solid phase is sodium cyanide dihydrate [25178-25-0], NaCH2H2O. The solubility of the dihydrate in grams of sodium cyanide per 100 grams of saturated solution is 26.01, at -15°C; 32.8, at 10°C; 34.2, at 15°C; and 45, at 34.7°C. The solubility of the anhydrous salt is less dependent on temperature.

Table 4. Physical Properties of Sodium Cyanide

Property	Value	
molecular weight	49.015	
melting point, °C	562	
boiling point, °C	1530	
density of solids, g mL		
cubic	1.60	
orthorhombic	1.62–1.624	
molten, at 700°C	~1.22	
density of solutions, at 25°C, g mL		
10° º NaCN	1.047	
20° º NaCN	1.098	
30° º NaCN	1.150	
vapor pressure, kPa ^a		
800°C	0.1013	
1360°C	41.8	
heat capacity ^b , 25–72°C, J (gK) ^c	1.40	
heat of fusion, J g ^c	179	
heat of formation, ΔH _f [°]	89.9 × 10 ³	
, NaCN(c), J mol ^c		
heat of vaporization, ΔH _{vap} , J g ^c	3041	
heat of solution, ^d ΔH _{soln} , J mol ^c	1548	
hydrolysis constant, K _a , 25°C	2.51 × 10 ⁻⁵	

^a To convert kPa to mm Hg, multiply by 7.5.

^b The heat capacity of sodium cyanide has been measured between 100 and 345 K (48).

^c To convert J to cal, divide by 4.184.

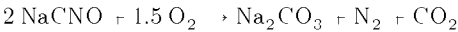
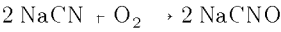
^d In 200 mol H₂O.

Sodium cyanide is soluble in liquid ammonia. At temperatures below -31°C, sodium cyanide pentaammoniate [69331-34-6], NaCN·5NH₃, separates in large flat crystals. At 15°C, 100 g anhydrous methanol dissolves 6.44 g anhydrous sodium cyanide; at 67.4°C, it dissolves 4.10 g. Sodium cyanide hemihydrate [69331-35-7], NaCN0.5 H₂O, has been obtained by recrystallization from cold 85° º alcohol. The system NaCN–NaOH–H₂O has been studied (48,49). Sodium cyanide is slightly soluble in formamide, ethanol, methanol, SO₂, furfural, and dimethylformamide.

Sodium chloride and sodium cyanide are isomorphous and form an uninterrupted series of mixed crystals. The ferrocyanide ion has a marked effect on the habit of sodium cyanide crystallized from aqueous solution (50). Sodium cyanide and sodium carbonate form a molten eutectic at approximately 53 wt % sodium carbonate and 465°C. The specific conductivity of molten 98° º sodium cyanide is 1.17 S cm (51).

Chemical Properties. When heated in a dry CO₂ atmosphere, sodium cyanide fuses without much decomposition. A brown-black color appears when water vapor and CO₂ are present at temperatures of 100°C below the fusion point. This color is presumably from the hydrogen cyanide polymer. Thermal dissociation of sodium cyanide has been studied in an atmosphere of helium at 600 to 1050°C and in an atmosphere of nitrogen at 1050 to 1255°C. It has been shown that the vapor phase over melt contains decomposition products (52).

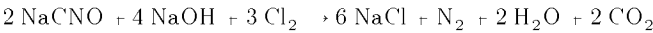
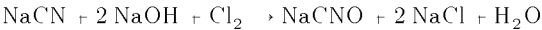
In the presence of a trace of iron or nickel oxide, rapid oxidation occurs when cyanide is heated in air, first to cyanate and then to carbonate:



Case hardening of steels using a sodium cyanide molten bath depends on these reactions where the active carbon and nitrogen are absorbed into the steel surface; hence the names carburizing and nitriding (see STEEL). The ease with which it is oxidized makes sodium cyanide a good reducing agent and the oxides of several metals, such as lead, tin, manganese, or copper, are readily reduced.

No reaction takes place below 500°C when sodium cyanide and sodium hydroxide are heated in the absence of water and oxygen. Above 500°C, sodium carbonate, sodium cyanamide [19981-17-0], sodium oxide, and hydrogen are produced. In the presence of small amounts of water at 500°C decomposition occurs with the formation of ammonia and sodium formate, and the latter is converted into sodium carbonate and hydrogen by the caustic soda. In the presence of excess oxygen, sodium carbonate, nitrogen, and water are produced (53).

Molten sodium cyanide reacts with strong oxidizing agents such as nitrates and chlorates with explosive violence. In aqueous solution, sodium cyanide is oxidized to sodium cyanate [917-61-3] by oxidizing agents such as potassium permanganate or hypochlorous acid. The reaction with chlorine in alkaline solution is the basis for the treatment of industrial cyanide waste liquors (45):

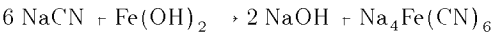


The pH value is usually maintained above 9 to avoid formation of nitrogen trichloride. At lower pH values, aqueous solutions react with chlorine to form cyanogen chloride (52).

Sodium cyanide, when fused with sulfur or a polysulfide, is converted into sodium thiocyanate [574-32-7]; this compound is also formed when a solution of sodium cyanide is boiled with sulfur or a polysulfide:

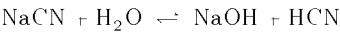


A solution of sodium cyanide shaken with freshly precipitated ferrous hydroxide is converted to a ferrocyanide:

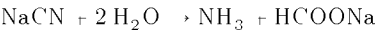


If the solution is acidified and a little ferric sulfate added, ferric ferrocyanide [14038-43-8], Fe₄[Fe(CN)₆]₃, is produced. This salt has a characteristic deep blue color, and the reaction may be used to test for the cyanide.

Aqueous solutions of sodium cyanide are slightly hydrolyzed at room temperature (54) according to the reversible reaction;

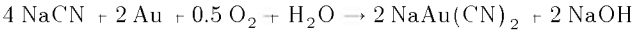


At temperatures above 50°C, irreversible hydrolysis to formate and ammonia becomes important. If the heat of reaction is not removed, the increased temperature accelerates the decomposition and can create high pressure in a closed vessel.



Hydrogen cyanide is a weak acid and can readily be displaced from a solution of sodium cyanide by weak mineral acids or by reaction with carbon dioxide, eg, from the atmosphere; however, the latter takes places at a slow rate.

In the presence of oxygen, aqueous sodium cyanide dissolves most metals in the finely divided state, with the exception of lead and platinum. This is the basis of the MacArthur process for the extraction of gold and silver from their ores that, in the case of gold, may be represented as follows:



The gold is then recovered from solution by precipitation with zinc dust, electrodeposit, or absorption on carbon. Sodium cyanide is used extensively in organic syntheses, especially in the preparation of nitriles (qv).

Manufacture

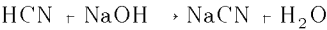
Alkali cyanides became commercially significant when the electroplating process for gold and silver was discovered about 1840. In 1899, the MacArthur-Forrest cyanide process, used for the extraction of gold and silver from low grade ores, was tested successfully on a commercial scale in New Zealand and South Africa. The MacArthur process was rapidly adopted in gold fields and resulted in a phenomenal growth in the cyanide manufacturing industry. World production of alkali cyanide rose from 5900 t yr in 1899 to 21,000 t yr by 1915. In the early 1990s the total world capacity was estimated to be in excess of 450,000 t. Annual usage in 1989 was about 340,000 t.

The historic Castner process that was developed in the 1800s and operated throughout the world until the late 1960s is now obsolete.



Lower yields and higher costs were the reasons for the obsolescence of this process.

Almost all sodium cyanide is manufactured by the neutralization or wet processes in which hydrogen cyanide reacts with sodium hydroxide solution.



Most plants use essentially purified anhydrous liquid hydrogen cyanide (sometimes vapors) to react with NaOH, generally fed as a 50° solution to the reactor. These raw materials must be sufficiently pure to yield a high quality, 99° sodium cyanide. Sodium hydroxide having low iron content is necessary to avoid interference with crystallization in the subsequent process steps. Purified hydrogen cyanide from the Andrussov or by-product hydrogen cyanide from acrylonitrile (qv) processes are used in most commercial sodium cyanide processes. Direct absorption of hydrogen cyanide in the product gases from Andrussov reactors into NaOH solution is practiced to some extent; however, the assay of the dry product is lower (96–97°) in this case owing to higher concentrations of sodium carbonate and sodium formate.

Most modern high tonnage plants use the reaction of high purity hydrogen cyanide and aqueous sodium hydroxide in a unit system that embodies evaporation of water and crystallization of the NaCN (55). Control of the system is critical to avoid hydrogen cyanide polymer formation, which would produce an off-white product; to minimize sodium formate formation, which would reduce the product purity; and to maximize the average crystal size. Figure 3 shows a typical sodium cyanide plant flow diagram.

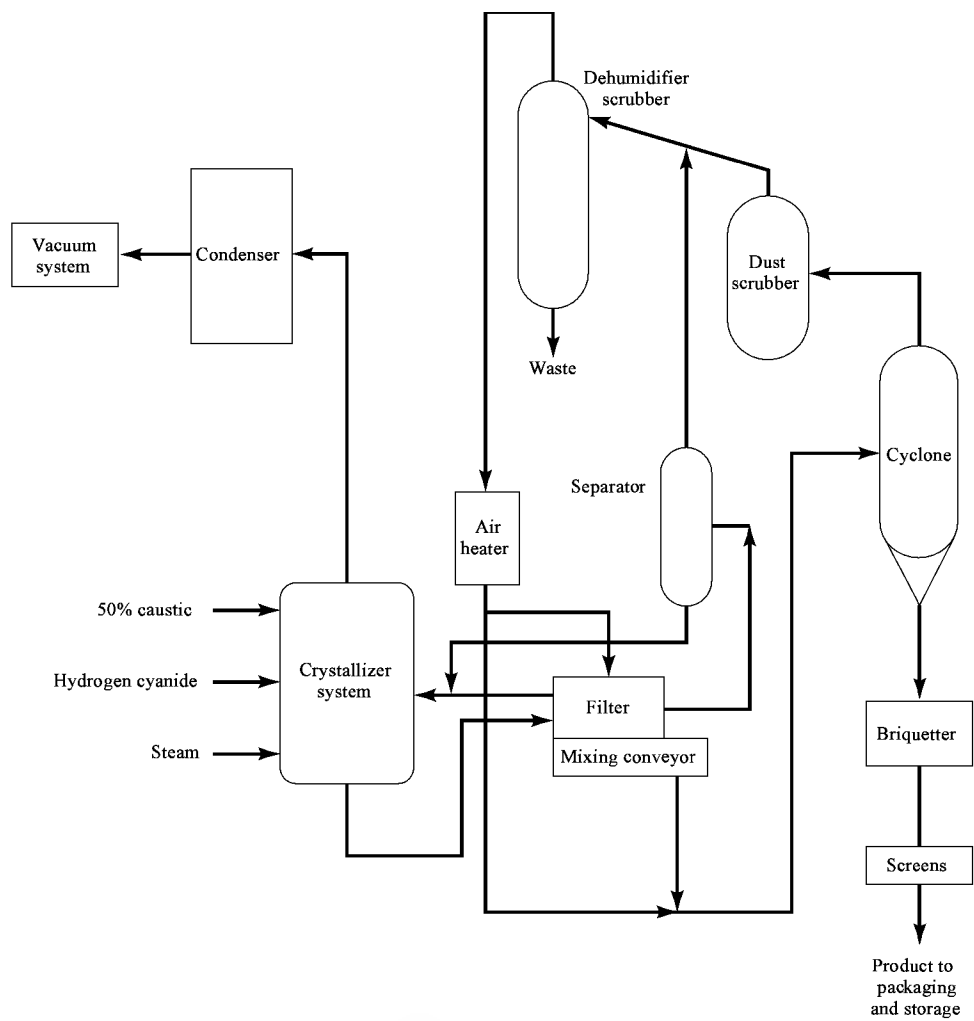


Fig. 3. Flow sheet for a sodium cyanide process.

A German process produces a high (99%) sodium cyanide assay by absorbing the gases from a BMA-type hydrogen cyanide reactor directly in sodium hydroxide solution (56). The resulting sodium cyanide solution is heated in a crystallizer to remove water, and form sodium cyanide crystals. The formation of larger crystals facilitates dewatering in the filtration step. In most modern processes, the moist salt from the filter is passed through a mixing conveyor to break up lumps. In some processes, air heated to 450°C is passed through the cake on the filter and through the mixing conveyor (57). Drying is completed in a hot-air conveyor-dryer (see DRYING). This type of adiabatic hot-air drying avoids overheating sodium cyanide when moisture is present, thereby minimizing the formation of sodium formate in the dried product. Inherently, sodium cyanide is a small (50 µm dia) crystal, that gives a dusty solid of low bulk density that must be compacted or fused into larger particles for safer handling. Melting the dry, dusty material and casting it into molds is rarely done because of the high energy requirement. Most modern processes employ mechanical compacting devices that produce either briquettes, granular products, or both (58). Briquettes or tablets are formed in approximate 15-g and 30-g sizes. The granular size is about 4 mm. In a typical process the finely divided dry crystals are compacted under heat and pressure in a roll press into briquettes having a density of 1.550 to 1.590. The briquettes are passed to a rotary screen where the fins, thin layers of material attached to the periphery of the briquette centerline, are removed and reprocessed. The finished briquettes pass into large storage bins from where they are loaded into rail-hopper cars or shipping bins or packaged into drums and other shipping containers. Plants for the production of sodium cyanide from Andrussow process or from acrylonitrile synthesis by-product hydrogen cyanide are operating in the United States, Italy, Japan, the UK, and Australia. In Germany, sodium cyanide is produced from BMA hydrogen cyanide, and in Australia one plant uses Fluohmic process hydrogen cyanide. The neutralization process is not energy intensive; added heat evaporates water formed in the reaction and water entering the system with the raw materials, which is 50% NaOH. The significant waste effluent contains 10–100 ppm NaCN and must be treated before disposal. A slight excess of NaOH must be maintained at all stages of processing to keep pH high and prevent forming color from black or brown polymer. The product as shipped must also contain a slight excess of NaOH so that clear, colorless sodium cyanide solutions result. Most sodium cyanide is sold in dry form to minimize the higher freight costs associated with shipping water. Appreciable tonnage is also sold as 30% aqueous solution.

Economic Aspects

Sodium cyanide is sold as granular or powder, pillow-shaped briquettes of 15-g and 30-g sizes, tablets of 30 g, and 30% aqueous solution. List price in 1991 was \$1.85 /kg. Typical analysis for the neutralization wet process product is given in Table 5. Sodium cyanide is packed in mild steel or fiber drums and in 1.4 t Flo-bins. Dry sodium cyanide is also shipped in wet-flo tank cars and trucks of up to 32 t net. At destination, water is circulated through the wet-flo car or trailer to dissolve the dry sodium cyanide at delivery. This type of shipment reduces freight costs and reduces environmental risks compared with 30% aqueous solution shipment. Safety regulations are imposed by the various shipping lines and by the countries in which cyanide is transported.

Table 5. Typical Analysis of Sodium Cyanide

Constituent	Concentration
NaCN, %	98–99
Na ₂ CO ₃ , %	0.3–0.9
HCOONa, %	0.4
NaOH, %	0.3
NaCl, %	0.01

S, ppm	<1.0
H ₂ O, ° o	0.06
Fe, ppm	10–20

Analysis

The cyanide content of a sodium cyanide sample may be determined by titration of an aqueous solution containing 10 g L with 0.1 *N* silver nitrate solution. Potassium iodide is the indicator. Use of the indicator *p*-dimethylaminobenzal–rhodamine [536-17-4], C₁₂H₁₂N₂OS₂, in the above procedure gives a more sensitive end point and can therefore be used for greater accuracy for low concentrations of cyanide. For still lower concentrations of cyanide (less than 0.2 ppm), the selective or specific ion electrode can be employed; this has the added advantage of functioning in colored or dirty solutions (see ELECTROANALYTICAL TECHNIQUES).

The sodium carbonate content may be determined on the same sample after a slight excess of silver nitrate has been added. An excess of barium chloride solution is added and, after the barium carbonate has settled, it is filtered, washed, and decomposed by boiling with an excess of standard hydrochloric acid. The excess of acid is then titrated with standard sodium hydroxide solution, using methyl red as indicator, and the sodium carbonate content is calculated.

Sulfide content is determined by titration with standard lead nitrate solution (1 g L). The titration is continued until a drop of the test solution on a filter paper ceases to produce a stain with a drop of lead nitrate solution.

Other ions, eg, ferrate, chloride, and formate, are determined by first removing the cyanide ion at ca pH 3.5 (methyl orange end point). Iron is titrated, using thioglycolic acid, and the optical density of the resulting pink solution is measured at 538 nm. Formate is oxidized by titration with mercuric chloride. The mercurous chloride produced is determined gravimetrically. Chloride ion is determined by a titration with 0.1 *N* silver nitrate. The end point is determined electrometrically.

Health and Safety Factors

Handling, storage, and the use of the alkali metal cyanides must be carried out by trained people. Most serious injuries and fatalities have been caused by inadvertently mixing these cyanides with acids, thereby releasing hydrogen cyanide. All necessary precautions should be taken to prevent cyanide salts from contacting liquid or airborne acids. Avoid storage near, or mixing with, acids. The present threshold limit value for 8-h exposure to cyanide dust is 5 mg m³ calculated as CN. Cyanide salts also must be protected from large concentrations of carbon dioxide to avoid hydrogen cyanide liberation. Carbon dioxide fire extinguishers should not be used. Cyanide salts as solids or solutions must be stored in tightly closed containers that must be protected from corrosion or damage. They should be stored so there is no contact with nitrate–nitrite mixtures or peroxides. Storage areas must be adequately ventilated and proper container labeling is mandatory.

Rubber gloves should be worn when handling dry salts. In addition, the following protective items should be used with solution or dusty salts: protective sleeves, aprons, shoes, boots or overshoes made of rubber, chemical safety goggles, full-face shield, and filter-type respirator (where dust is present). Cyanide spills should be flushed to a contained area where treatment to destroy the cyanide can be carried out. Avoid flushing to drains or ditches that may contain acids. In the event that cyanide salts or solutions contact the eyes, they should be flushed for 15 min with a copious, gentle flow of water followed by immediate medical attention. Eating, smoking, and chewing should be forbidden in areas where cyanide salts are handled. Employees should be required to wash carefully after working with cyanide salts and before eating, smoking, or chewing (59).

Uses

At one time, electroplating was the largest single use for sodium cyanide, especially for zinc (60), copper, brass, and cadmium. Gold, gold alloys, and silver are also plated onto base metals from sodium cyanide solutions. Substantial decline in cyanide electroplating has occurred, however, owing to tighter restrictions on cyanide discharge and conservation of plating and rinse solutions.

Use for heat treating is also small compared to the past. It has been supplanted to a large extent by furnaces using special atmospheric conditions. Heat treatment salts containing sodium cyanide are used for small metal parts when selective case hardening is required.

The principal use of sodium cyanide is for recovery of precious metals. Recovery of gold by cyanidation is the largest single mining use for sodium cyanide and has been growing owing to high gold prices (61). The sodium cyanide, in dilute solution and in the presence of oxygen, is used to dissolve the gold out of the ore. The water-soluble sodium dicyanoaurated(I) [15280-09-8], NaAu(CN)₂, is formed. From this dilute solution, a typical process adsorbs the gold complex onto carbon. Then stronger cyanide solutions are used to elute the gold off of the carbon for electrochemical recovery from this solution. Metallic silver frequently occurring with the gold is recovered at the same time.

Chemical uses for sodium cyanide may constitute as much as 5° o of the total and are in five general categories: dyes, including optical brighteners (qv); agricultural chemicals; pharmaceuticals; chelating or sequestering agents (see CHELATING AGENTS); and specialties. Sodium cyanide is used in the preparation of nitriles (qv), carbylamines (isonitriles), cyano fatty acids, and heavy metal cyanides (see also CYANOCARBONS). Sodium cyanide reacts with dextrose hydrate in aqueous solution to produce glucoheptonate [31138-65-5] (62). Phenylglycine [109-1-5], the intermediate in the manufacture of indigo [482-89-3], is manufactured using sodium cyanide. Substantial quantities of sodium cyanide are used in sodium ferrocyanide [13601-19-9], Na₄ Fe(CN)₆, for the manufacture of Prussian blue. Sodium cyanide solution treated with FeCl₂ in the presence of alkali metal hydroxide with careful adjustment of pH produces yellow Na₄Fe(CN)₆ (63). Sodium cyanide is used to produce sodium nitrilotriacetate [10042-84-9], SNTA, which is an increasingly popular chelating agent as well as a phosphate replacement in certain areas of the world. Cyanuric chloride [108-77-0] is produced in small quantities from sodium cyanide (see CYANURIC AND ISOCYANURIC ACIDS).

Miscellaneous uses for sodium cyanide include heat treating, metal stripping, and compounds used for clearing smut. Treatment of wood chips with sodium cyanide and CaCl₂ reportedly increases the kraft cooking yield of pulp (qv) (64).

POTASSIUM CYANIDE

Potassium cyanide [151-50-8], KCN, a white crystalline, deliquescent solid, was initially used as a flux, and later for electroplating, which is the single greatest use in the 1990s. The demand for potassium cyanide was met by the ferrocyanide process until the latter part of the nineteenth century, when the extraordinary demands of the gold mining industry for alkali cyanide resulted in the development of direct synthesis processes. When cheaper sodium cyanide became available, potassium cyanide was displaced in many uses. With the decline in the use of alkali cyanides for plating the demand for potassium cyanide continues to decline. The total world production in 1990 was estimated at about 4500 t, down from 7300 t in 1976.

Commercial potassium cyanide made by the neutralization or wet process contains 99° o KCN; the principal impurities are potassium carbonate, formate, and hydroxide. To prepare 99.5 + % KCN, high quality hydrogen cyanide and KOH must be used.

Properties

Physical Properties. The physical properties of potassium cyanide are given in Table 6. Unlike sodium cyanide, potassium cyanide does not form a dihydrate.

Table 6. Physical Properties of Potassium Cyanide

Property	Value
molecular weight	65.11
melting point, °C	634
density, g mL	
cubic	1.55
orthorhombic at 60°C	1.62
specific heat, 25–72°C, J g ^a	1.01
heat of fusion, J mol ^a	14.7 × 10 ³
heat of formation, Δ <i>H</i> _{<i>f</i>} [°] , J mol ^a	113 × 10 ³
heat of solution, Δ <i>H</i> _{soln} , J mol ^a	12550
hydrolysis constant, 25°C	2.54 × 10 ^{−5}
solubility at 25°C, g 100 g H ₂ O	71.6
resistivity, Ωcm	
0.25 <i>N</i> soln	70
0.5 <i>N</i> soln	15
1.0 <i>N</i> soln	10
2.0 <i>N</i> soln	5

^a To convert J to cal, divide by 4.184.

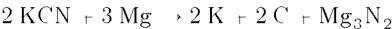
The solubility of potassium cyanide in nonaqueous solvents is as follows: in anhydrous liquid ammonia, 4.55 g 100 g NH₃ at −33.9°C; 4.91 g 100 g methanol at 19.5°C; 0.57 g 100 g ethanol at 19.5°C; 146 g L solution in formamide at 25°C; 41 g 100 g hydroxylamine at 17.5°C; 24.24 g 100 g glycerol of specific gravity 1.2561 at 15.5°C; 0.73 g L solution in phosphorus oxychloride at 20°C; 0.017 g 100 g liquid sulfur dioxide at 0°C; and 0.22 g 100 g dimethylformamide at 25°C.

At room temperature, potassium cyanide has fcc crystal structure (65). Average potassium cyanide crystals inherently have three to four times the mass or size of sodium cyanide crystals.

Chemical Properties. Potassium cyanide is readily oxidized to potassium cyanate [590-28-3] by heating in the presence of oxygen or easily reduced oxides, such as those of lead or tin or manganese dioxide, and in aqueous solution by reaction with hypochlorites or hydrogen peroxide.

Dry potassium cyanide in sealed containers is stable for many years. An aqueous solution of potassium cyanide is slowly converted to ammonia and potassium formate; the decomposition rate accelerates with increasing temperature. However, at comparable temperatures the rate of conversion is far lower than that for sodium cyanide; only about 25% as great.

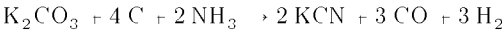
Many reactions can be carried out between potassium cyanide and organic compounds with the alkalinity of the KCN acting as a catalyst; these reactions are analogous to reactions of sodium cyanide. The reactions of potassium cyanide with sulfur and sulfur compounds are also analogous to those of sodium cyanide. Potassium cyanide is reduced to potassium metal and carbon by heating it out of contact with air in the presence of powdered magnesium. Magnesium is converted to the nitride:



Beryllium, calcium, boron, and aluminum act in a similar manner. Malonic acid is made from monochloroacetic acid by reaction with potassium cyanide followed by hydrolysis. The acid and the intermediate cyanoacetic acid are used for the synthesis of polymethine dyes, synthetic caffeine, and for the manufacture of diethyl malonate, which is used in the synthesis of barbiturates. Most metals dissolve in aqueous potassium cyanide solutions in the presence of oxygen to form complex cyanides (see COORDINATION COMPOUNDS).

Manufacture

Potassium cyanide was made by the Beilby process before the introduction of the neutralization or wet process. In the Beilby process, cyanide is made according to the following overall reaction:



In this dry process, ammonia gas passes into a molten mixture of potassium carbonate and charcoal. Although purity of the product is high, this process became obsolete because of the lower costs of the neutralization process.

Potassium cyanide is manufactured by the reaction of an aqueous solution of potassium hydroxide and hydrogen cyanide:



The cyanide, which crystallizes in the anhydrous state from aqueous solution, is recovered by evaporation under reduced pressure, filtration, and drying. Because the crystal size is significantly larger than sodium cyanide it can be sold in powder form without excessive dusting. However, it tends to cake in the shipping container and is often compacted and granulated to larger sizes.

Uses

Potassium cyanide is primarily used for fine silver plating but is also used for dyes and specialty products (see ELECTROPLATING). Electrolytic refining of platinum is carried out in fused potassium cyanide baths, in which a separation from silver is effected. Potassium cyanide is also a component of the electrolyte for the analytical separation of gold, silver, and copper from platinum. It is used with sodium cyanide for nitriding steel and also in mixtures for metal coloring by chemical or electrolytic processes.

Potassium cyanide like sodium cyanide, is shipped in steel or fiber drums. Potassium cyanide costs more than sodium cyanide, primarily because of the higher price of potassium hydroxide. The 1991 price was \$3.88/kg.

OTHER CYANIDES

Lithium, Rubidium, and Cesium Cyanides

Lithium cyanide [2408-36-8], rubidium cyanide [19073-56-4], and cesium cyanide [21159-32-0] are white or colorless salts, isomorphous with potassium cyanide. In physical and chemical properties these cyanides closely resemble sodium and potassium cyanide. As of this writing these cyanides have no industrial uses.

All of these alkali metal cyanides may be prepared by passing hydrogen cyanide into an aqueous solution of the hydroxide or by precipitating a

solution of barium cyanide [542-62-1] with lithium, rubidium, or cesium sulfate. A product with fewer contaminants may be obtained by the reaction of the base in absolute alcohol or dry ether with anhydrous hydrogen cyanide (66). In another method of preparation, a suspension of rubidium, cesium, or lithium metals in anhydrous benzene is treated with anhydrous hydrogen cyanide, and the benzene subsequently removed by evaporation under reduced pressure.

These cyanides are all soluble in water. The cyanide ion is weakly held so that water solutions have a much stronger odor of hydrogen cyanide above them than sodium and potassium cyanide solution.

Lithium cyanide melts at 160°C. In the fused state the specific gravity at 18°C is 1.075. It is highly hygroscopic. Rubidium cyanide is not hygroscopic and is insoluble in alcohol or ether. Cesium cyanide is highly hygroscopic.

Lithium cyanide decomposes to cyanamide and carbon below about 600°C. This decomposition is similar to the alkaline-earth cyanides (67). Iron accelerates decomposition and, when heated with 10% iron at 500°C for 15 h, lithium cyanide is completely converted to lithium cyanamide [2408-36-8].

Ammonium Cyanide

Ammonium cyanide [12211-52-8], NH₄CN, a colorless crystalline solid, is relatively unstable, and decomposes into ammonia and hydrogen cyanide at 36°C. Ammonium cyanide reacts with ketones (qv) to yield aminonitriles. Reaction of ammonium cyanide with glyoxal produces glycine. Because of its unstable nature, ammonium cyanide is not shipped or sold commercially. Unless it is kept cool and dry, decomposition releases vapors and forms black hydrogen cyanide polymer.

Ammonium cyanide may be prepared in solution by passing hydrogen cyanide into aqueous ammonia at low temperatures. It may also be prepared from barium cyanide and ammonium sulfate, or calcium cyanide with ammonium carbonate. It may be prepared in the dry state by gently heating a mixture of potassium cyanide or ferrocyanide and ammonium chloride, and condensing the vapor in a cooled receiver. Ammonium cyanide is soluble in water or alcohol. The vapor above solid NH₄CN contains free NH₃ and HCN, a very toxic mixture.

Calcium Cyanide

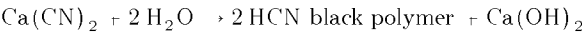
Crude calcium cyanide [592-01-8], about 48 to 50 eq % sodium cyanide, is the only commercially important alkaline-earth metal cyanide, and output tonnage has been greatly reduced. This product, commonly called black cyanide, is marketed in flake form as a powder or as cast blocks under the trademarks Aero and Cyanogas of the American Cyanamid Co.

The calcium cyanamide process was first discovered in Germany in the early 1900s during experimental attempts to produce cyanide cheaply (68,69). Following the discovery, the best attempts resulted in only 17.5% sodium cyanide equivalent product. Because of a shortage of cyanide in the United States in 1916, the American Cyanamid Co. developed a batch process giving higher yields that evolved into a continuous electric-furnace process in 1919. Subsequent improvements increased the strength to 48 to 50% sodium cyanide equivalent. Calcium cyanide solutions of about 15% concentration have been prepared for use in gold extraction in South Africa from hydrogen cyanide and lime.

Physical and Chemical Properties. Because of decomposition, the melting point of calcium cyanide can only be estimated by extrapolation to be 640°C (70).

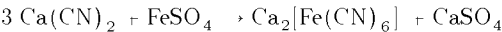
Calcium cyanide diammoniate [69365-88-4], Ca(CN)₂·2NH₃, is formed in liquid ammonia by reaction of calcium hydroxide or nitrate with ammonium cyanide. Deammoniation under heat and high vacuum yields calcium cyanide, a white powder, which is readily hydrolyzed to hydrogen cyanide. The diammoniate is more stable than the cyanide containing no ammonia. Calcium cyanide can be made by passing ammonia and hydrogen cyanide over calcium hydroxide. A compound having even higher cyanide content is obtained by subjecting calcium carbide (see CARBIDES) to the action of liquid hydrogen cyanide in the presence of 0.5–5% of water; it has also been formed in a nonaqueous medium such as dimethylformamide. These processes are distinct from those for the fused cyanides in which nitrification of calcium carbide yields calcium cyanamide [156-62-7], which in turn produces calcium cyanide by reaction with carbon.

Aqueous solutions of calcium cyanide prepared even at low temperature turn yellow or brown owing to the formation of HCN polymer. Calcium cyanide hydrolyzes readily.

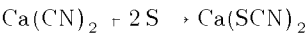


The presence of brown polymer in solid form is sometimes noted even in dry calcium cyanide that has been stored for long periods. Calcium cyanide is decomposed by carbon dioxide, acids, and acidic salts liberating hydrogen cyanide.

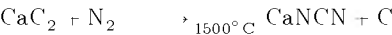
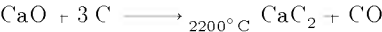
Ferrocyanides are produced by reaction of ferrous salts



With sulfur in aqueous medium, calcium cyanide forms calcium thiocyanate [2092-16-2]



Manufacture. Calcium cyanide is made commercially from lime [1305-78-3], CaO, coke, and nitrogen. The reactions are carried out in an electric furnace (69).



The resulting melt is cooled rapidly to prevent reversion to calcium cyanamide. The product is marketed in the form of flakes, dark gray because of the presence of carbon. Typical composition is shown in Table 7. Because the rate of hydrogen cyanide evolution is relatively high, it is readily adaptable to fumigation. Specific gravity of the product is 1.8 to 1.9. The price of black cyanide is generally lower than sodium cyanide; it is manufactured in Canada and South Africa.

Table 7. Typical Analysis of Calcium Cyanide

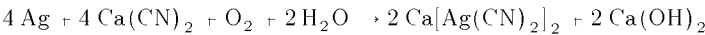
Constituent	Concentration, wt %	
Ca(CN) ₂	45–50	
NaCl	32	
CaO	12	
graphite	1–2	
CaC ₂	2–3	
CaNCN	2.5	
SiO ₃	0.9	
M ₂ O ₃ + Al ₂ O ₃	2.3 ^a	

CaS	0.4
CaF	0.5
MgO	0.3

^a M may be Fe, Cr, or less commonly Ni, Co, or other metals.

Safety Precautions. Precautions similar to those used for sodium cyanide should be used for black cyanide. Contact with acidic compounds generates hydrogen cyanide gas. Some additional precaution must be taken to minimize contact with water or even the humidity of the atmosphere because decomposition and hydrogen cyanide evolution occurs. Black cyanide is shipped in steel drums and is designated as a Class B poison. It is extremely toxic to humans, animals, and fish. Ingestion, contact with the skin, inhalation of the dust or hydrogen cyanide evolved should be avoided. Black cyanide should be stored in tight containers under moisture-free conditions; the small amount of contained carbide can evolve acetylene. Adequate ventilation and protective equipment similar to that for sodium cyanide should be provided in handling the solid cyanide and in the preparation of aqueous solutions.

Uses. The extraction or cyanidation of precious metal ores was the first, and is still the largest, use for black cyanide (71). The leaching action of the cyanide results from the formation of soluble cyanide complexes.



This reaction also applies to gold.

The flake cyanide is dissolved and added to suspensions of finely ground ore under agitation and in the presence of air. The dissolved gold and silver are precipitated by the addition of zinc dust and the crude precipitate is refined to produce bullion.

Black cyanide is also used in the processes in which the gold complexes are adsorbed on carbon. To recover the gold from the carbon a more concentrated solution of sodium cyanide is used.

The use of black cyanide as a fumigant and rodenticide makes use of the atmospheric humidity action that liberates hydrogen cyanide gas. It can only be used effectively in confined spaces where hydrogen cyanide builds up to lethal concentrations for the particular application. Black cyanide is also used in limited quantities in the production of prussiates or ferrocyanides (see IRON COMPOUNDS).

Other Alkaline-Earth Cyanides

Magnesium cyanide [22400-99-3], Mg(CN)₂, may be formed by reaction of finely divided magnesium metal and ammonium cyanide in liquid ammonia. Removal of the ammonia yields a white solid, magnesium cyanide diammoniate [69309-45-1]; deammoniation by heating under vacuum gives the cyanide, which is similar in properties to calcium cyanide. Strontium cyanide [60448-24-0] and barium cyanide [60448-23-9] are somewhat more stable than calcium and magnesium cyanides and their preparation is by reaction of strontium or barium hydroxide and hydrogen cyanide in water followed by low temperature evaporation under vacuum. The cyanides of magnesium, barium, and strontium have not been marketed.

BIBLIOGRAPHY

"Cyanides and Cyanogen Compounds" in *ECT* 1st ed., Vol. 4, pp. 689–712, by W. J. Miller and D. E. Nye, E.I. du Pont de Nemours & Co., Inc., and R. B. Mooney, American Cyanamid Co., "Hydrogen Cyanide" under "Cyanides" in *ECT* 2nd ed., Vol. 6, pp. 574–585, by P. D. Montgomery, Monsanto Co.; "Cyanides (Alkali Metal)" under "Cyanides" in *ECT* 2nd ed., Vol. 6, pp. 585–601, by R. B. Mooney and J. P. Quin, Imperial Chemical Industries Ltd; "Alkaline-Earth Metal Cyanides" under "Cyanides" in *ECT* 2nd ed., Vol. 6, pp. 601–704, by R. B. Booth, American Cyanamid Co.; "Cyanides" in *ECT* 3rd ed., Vol. 7, pp. 307–334, by W. R. Jenks, E.I. du Pont de Nemours & Co., Inc.

1. C. Ponnampерuma, *Sci. J.* **1**(3), 39045 (1965).
2. J. P. Ferris *Science* **203**, 1135 (1979); C. N. Matthews, *Science* **203**, 1136 (1979).
3. R. J. Cicerone and R. Zellner, *J. Geophys. Res.* **88**(C15), 10689 (1983).
4. A. W. Cobb and J. H. Walton, *J. Chem. Phys.* **41**, 351 (1957).
5. S. Barrett and A. F. Titley, *J. Chem. Soc.* **115**, 902 (1919).
6. A. Bernthsen, *Ann.* **192**, 1 (1878).
7. **U.S. Pat. 2,712,493** (July 5, 1955), W. Moje (to E.I. du Pont de Nemours & Co., Inc.).
8. **Ger. Pat. 1,056,101** (Apr. 30, 1959), H. Zima (to Rohm & Haas GmbH).
9. **U.S. Pat. 2,672,398** (Mar. 16, 1954), H. Huemer, H. Schulz, and W. Pohl (to Deutsche Gold und Silber-Scheideanstalt).
10. R. Oleinek, *Rev. Chim.* **16**, 19 (1965).
11. **U.S. Pat. 2,399,361** (Apr. 30, 1946), B. S. Lacy and W. S. Hinegardner (to E.I. du Pont de Nemours & Co., Inc.).
12. **U.S. Pat. 2,762,798** (Sept. 11, 1956), N. L. Hardwicke and G. W. Walpert (to Monsanto Co.).
13. **U.S. Pat. 2,434,606** (Jan. 13, 1948), E. L. Carpenter (to American Cyanamid Co.).
14. G. Schlesinger and S. L. Miller, *J. Am. Chem. Soc.* **95**, 3729 (1973).
15. L. I. Krimen and J. Donald, *Org. React.* **17**, 213 (1969).
16. **Can. Pat. 902,099** (Apr. 26, 1971), H. Neumaier and co-workers (to Knapsack Aktiengesellschaft).
17. **Ger. Pat., 2,205,983** (Aug. 16, 1973), W. Gruber and co-workers (to Ruhm GmbH).
18. **U.S. Pat. 3,177,215** (Apr. 6, 1965), R. W. Foreman and co-workers (to Standard Oil Co.).
19. **U.S. Pat. 3,988,360** (Oct. 26, 1976), R. Gaudette and J. Philbrook (to W. R. Grace & Co.).
20. R. C. Sheridan and E. H. Brown, *J. Org. Chem* **30**(2), 668 (1965).
21. **Jpn. Pat. 7,404,207** (Jan. 31, 1974), Y. Hamamoto (to Kyowa Gas Chemical Industry Co., Ltd.).
22. **U.S. Pat. 3,666,787** (May 30, 1972), O. Webster (to E.I. du Pont de Nemours & Co., Inc.).
23. **U.S. Pat. 1,934,839** (Nov. 14, 1933), L. Andrussow (to I. G. Farbenindustrie, AG).
24. L. Andrussow, *Agnew. Chem.* **48**, 593 (1935).
25. L. Andrussow, *Bull. Soc. Chim. Fr.* **18**, 45 (1951).
26. L. Andrussow, *Chem. Ing. Tech.* **27**, 469 (1955).
27. L. Andrussow, *Chim. Ind. Genie Chim.* **86**, 39 (1961).
28. C. T. Kautter and W. Leitberger, *Chem. Ing. Tech.* **25**, 599 (1953).
29. P. W. Sherwood, *Pet. Eng.* **31**, C-22, C-51 (1959).
30. **U.S. Pat. 3,215,495** (Nov. 2, 1965), W. R. Jenks and A. W. Andresen (to E.I. du Pont de Nemours & Co., Inc.).
31. **U.S. Pat. 3,360,355** (Dec. 26, 1967), W. R. Jenks (to E.I. du Pont de Nemours & Co., Inc.).
32. **U.S. Pat. 3,104,095** (Sept. 24, 1963), W. R. Jenks and R. M. Shepherd (to E.I. du Pont de Nemours & Co., Inc.).
33. R. Rodiger, *Chem. Tech. (Berlin)* **10**, 135 (1958).
34. F. Endter, *Chem. Ing. Tech.* **30**, 305 (1958).
35. *Chem. Week* **83**(4), 70 (1958).
36. **Ger. Pat. 2,014,523** (Apr. 17, 1970) (to Degussa).
37. N. B. Shine, *Chem. Engr. Prog.* **67**, 2 (1971).
38. G. E. Johnson and co-workers, *U.S. Bureau of Mines Report*, No. RI6994, Washington, D.C., 1967.

39. **Ital. Pat. 845,992** (July 2, 1969) (to Montecatini Edison).

40. R. J. H. Voothoeve and co-workers, *J. Catal.* **45**, 297 (1976).

41. **Ger. Pat. 1,143,497** (Feb. 14, 1963), G. Schulze and G. Weiss (to Badische Anilin-und Soda-Fabrik AG).

42. D. Hasenberg, *HCN Synthesis on Polycrystalline Platinum and Rhodium*, dissertation, University of Minnesota, 1984.

43. **U.S. Pat. 2,797,148** (June 25, 1957), H. C. Carlson (to E.I. du Pont de Nemours & Co., Inc.).

44. E. Koberstein, *Ind. Eng. Chem. Process Des. Dev.* **12**(4) (1973).

45. H. Roos and M. Schmidt, *Chem. Ing. Tech.* **49**, 369 (1977).

46. J. L. Huiatt and co-workers, eds., *Proceedings of a Workshop—Cyanide From Mineral Processing*, University of Utah, Salt Lake City, 1982.

47. T. Green, *The Prospect for Gold*, Walker Publishing Co. Inc., 1987.

48. E. D. Oliver and S. E. J. Johnson, *J. Am. Chem. Soc.* **76**, 4721 (1954).

49. V. A. Kireev and L. I. Vogranskaya, *J. Gen. Chem. USSR* **5**, 963 (1935).

50. **U.S. Pat. 2,949,341** (Aug. 16, 1960), C. P. Green (to E.I. du Pont de Nemours & Co., Inc.).

51. E. Rishkevich, *Z. Electrochem.* **39**, 531 (1933).

52. E. W. Guernsey and M. S. Sherman, *J. Am. Chem. Soc.* **48**, 695 (1926).

53. R. Holtze, *A. Anorg. Allgem. Chem.* **214**, 65 (1933).

54. B. Ricca and G. D'Amore, *Gazz. Chim. Ital.* **79**, 308 (1949).

55. **U.S. Pat. 2,993,754** (July 25, 1961), W. R. Jenks and J. S. Linder (to E.I. du Pont de Nemours & Co., Inc.).

56. **U.S. Pat. 3,619,132** (Nov. 9, 1971), H. J. Mann and co-workers (to Degussa).

57. **U.S. Pat. 2,944,344** (July 12, 1960), C. P. Green and W. R. Jenks (to E.I. du Pont de Nemours & Co., Inc.).

58. **U.S. Pat. 3,615,176** (Oct. 26, 1971), W. R. Jenks and O. W. Shannon (to E.I. du Pont de Nemours & Co., Inc.).

59. *Sodium Cyanide Material Safety Data Sheet*, E.I. du Pont de Nemours & Co., Inc., Wilmington, Del., 1990.

60. H. Geduld, *Metal Finish* **74**(1), 45 (1976).

61. M. C. Fuerstenau and J. L. Hendrix, eds., *Advances in Gold and Silver Processing Proceedings*, Society for Mining, Metallurgy, & Exploration, Inc., Littleton, Colo., 1990.

62. **U.S. Pat. 3,679,659**, (July 25, 1972), C. Zak (to Belzak Corp.).

63. **U.S. Pat. 3,695,833** (Oct. 3, 1972), O. Wiedeman and co-workers (to American Cyanamid Co.).

64. **U.S. Pat. 3,883,391** (May 13, 1975), R. B. Phillips (to International Paper Co.).

65. R. M. Bozorth, *J. Am. Chem. Soc.* **44**, 317 (1922).

66. J. Meyer, *Z. Anorg. Allgem. Chem.* **115**, 203 (1920).

67. A. Perret and J. Riethmann, *Helv. Chim Acta* **26**, 740 (1943).

68. **U.S. Pat. 776,314** (Nov. 29, 1904), A. Frank.

69. *Mining Chemicals Handbook*, American Cyanamid, Wayne, N.J., 1986.

70. G. Peterson and H. H. Franck, *Z. Anorg. Allgem. Chem.* **237**, 1 (1938).

71. J. V. N. Dorr and F. L. Bosqui, *Cyanidation and Concentration of Gold and Silver Ores*, McGraw-Hill Book Co., Inc., New York, 1950.

Lawrence D. Pesce
E. I. du Pont de Nemours & Co., Inc.

CYANINE DYES

This large class of dyes with a wide variety of colors shows absorptions that cover the ultraviolet to the infrared region and, as a group, cover a wider span of the spectrum than those of any other dye class. The cyanine dyes are also among the oldest known class of synthetic dyes; the first dye was discovered in 1856 (1). The name cyanine (from the Greek *kyanos*) was attributed to its beautiful blue color; however, the dye was extremely fugitive to light and of no practical use at the time for ordinary (fabric) dyeing purposes. The great usefulness of cyanines was discovered later in photography, and they include the most powerful photographic sensitizing dyes known (2–4) (see COLOR PHOTOGRAPHY). There are several important reasons for the cyanines' prominence as sensitizers: first, they have high light absorption per molecule coupled in many cases with a single absorption band in the visible or infrared spectral region, which gives very color-selective absorption of light; second, they have a tendency to form dye aggregates that have even narrower, more color-selective absorptions than the monomeric dyes themselves; and third, they have high chemical and photochemical reactivity for dyes adsorbed to silver halides, which leads to efficient participation in photographic sensitization processes (see DYES, SENSITIZING).

The cyanine class of dyes is also useful in biological, medical, laser, and electro-optic applications. Dyes marketed as Povon [3546-41-6] (5) and Dithiazanine [7187-55-5] (6) are useful anthelmintics, and Indocyanine Green [3599-32-4] (7) is an infrared-absorbing tracer for blood-dilution medical diagnoses. "Stains-All" is a well-studied biological stain (8) and Merocyanine 540's photochemotherapeutic activity is known in some detail (9). Many commercially available red and infrared laser dyes are cyanines (10).

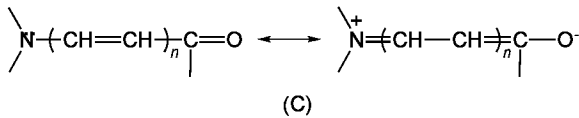
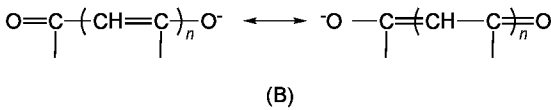
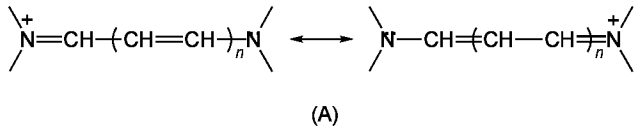
The cyanine dye literature has been reviewed (2,11–19). Ref. 12 is the best source for the heterocyclic chemistry of cyanine and related dyes. Several new syntheses and many new dye structures are discussed in later reviews (13,16) combined with compilations of physical and photophysical data (16,17).

Well-developed synthetic methods allow cost-effective manufacture of cyanines for commercial applications as well as a high degree of dye-structure design for new and innovative studies (20,21), such as solar cells, electrophotography, Langmuir-Blodgett nonlinear optical layers, and photoreceptors for processes activated by infrared solid-state lasers. Thus, tailoring the many characteristics of a dye is a well-practiced art in the cyanine class. Combinations of heterocycles, substituents, and chromophore lengths can yield a series of dye structures having parallel shifts in oxidation-reduction potentials almost independent of absorption wavelength. Steric features of substituents either enhance or decrease aggregation. Solubility in either aqueous or hydrocarbon solvents can be provided by other substituents, almost independently from those that change redox potentials or steric properties. Controlling the number of conformations of the polymethine chain, achieved by several synthetic routes, is important to enhanced infrared absorption strength.

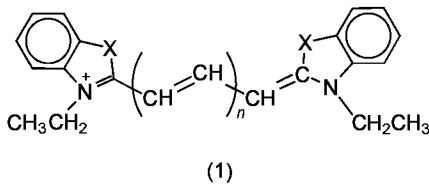
Color and Constitution

The color and constitution of cyanine dyes may be understood through detailed consideration of their component parts, ie, chromophoric systems, terminal groups, and solvent sensitivity of the dyes. Resonance theories have been developed to accommodate significant trends very successfully. For an experienced dye chemist, these are useful in the design of dyes with a specified color, band shape, or solvent sensitivity. More recently, quantitative values for reversible oxidation-reduction potentials have allowed more complete correlation of these dye properties with organic substituent constants.

Chromophoric Systems. The primary types of chromophores for cyanine dyes are the amidinium-ion system (A), the carboxyl-ion system (B), and the dipolar amidic system (C). For each system two extreme resonance structures are shown in (A), (B), and (C) where the formal charges are located at the ends of the chromophore. Intermediate resonance structures, with the charges closer to the center of the chromophore or with additional dipoles, are less important in the resonance picture of dyes. However, structural changes that favor intermediate forms have significant effects on the color of symmetrical dyes containing (A). For the amidic dyes (C), structural features stabilizing both neutral and dipolar extreme resonance forms in an equivalent manner give dyes absorbing at longer wavelengths.



The important characteristics that influence the absorption wavelengths for these dyes are the length of the conjugated chain and the nature of the terminal group. Many of the early cyanine dyes comprised a chain with an odd number of methine carbon atoms (—CH—) and two heterocycles, eg, quinoline or benzothiazole. Historically, the terms simple cyanine, carbocyanine, dicarbocyanine, and such were used to designate both the specific dyes derived from quinoline as well as generic dye structures (from other heterocycles) with one, three, five, etc, methine carbon atoms (see structure (1)). In the dyes from quinoline, which can also be designated quinocyanines, the ring position attached to the methine chain and the N-substituent are usually specified. For example, 1,1'-diethyl-2,2'-cyanine (or 1,1'-diethyl-2,2'-quinocyanine) is dye (1), $\text{X} = \text{CH} = \text{CH}$, and $n = 0$. Pinacyanol is dye (1), $n = 1$ and can be named as 1,1'-diethyl-2,2'-carbocyanine iodide, or as 1,1'-diethyl-2,2'-quinocarbocyanine iodide. Chemical Abstracts names pinacyanol as a substituted quinolinium iodide, ie, quinolinium, 1-ethyl-2-[3-(1-ethyl-2(1H)-quinolyldiene)-1-propenyl]-iodide.



Dyes derived from the primary chromophores (B) and (C) are designated oxonols and merocyanines, although the term neutrocyanine has also been used for (C). For merocyanines the simple merocyanine designation refers to a dye with zero methine carbon atoms ($n = 0$) (ie, the shortest possible linkage that retains the chromophore between the terminal groups); the term merocarbocyanine designates dyes having chromophore (C) with two methine carbon atoms ($n = 1$).

Several examples of the chromophoric systems (A), (B), and (C) are shown in Figure 1. The early dyes were single chromophore structures of the type (A): Williams' cyanine [862-57-7], Pinacyanol [605-91-4], and thiocarbocyanine [905-97-5]. The more complicated dye structures in Figure 1 still contain these chromophoric systems.

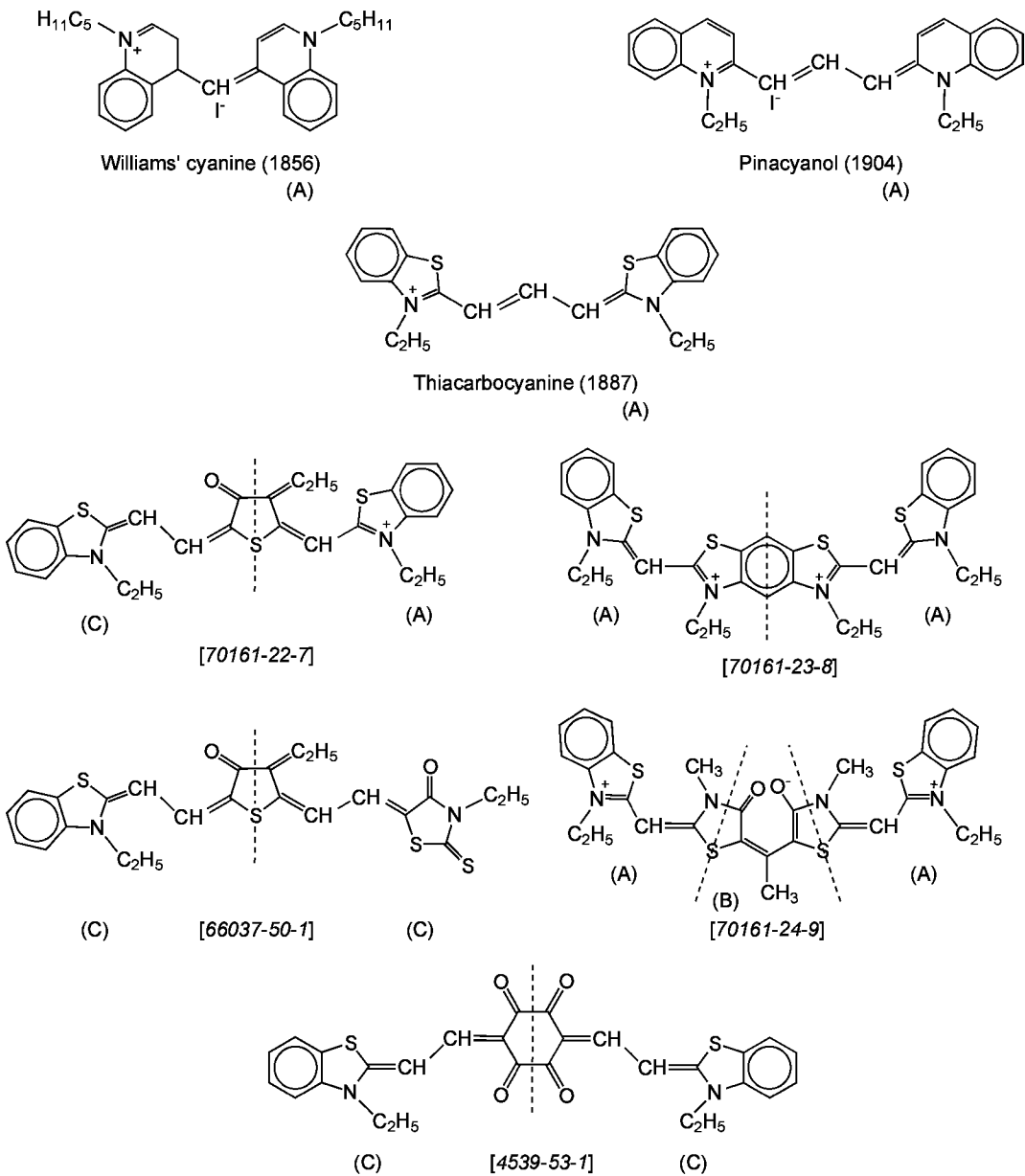


Fig. 1. Chromophoric systems. Note that (A) = amidinium-ion system (a cyanine), (B) = carboxyl-ion system (an oxonol), and (C) = dipolar amidic system (a merocyanine). Counterions are usually not specified in the text.

Dyes that differ only by the number of vinyl groups ($-\text{CH}=\text{CH}-$) in the methine chain are termed a vinylogous series. Absorption maxima for vinylogous series of dyes like (2) to (6) in Table 1 shift to longer wavelengths as the methine chain length increases. With the usual exception of the first member of each series, the shift approximates 100 nm per vinyl group in most symmetric chromophores like (2) (amidinium-ion system, A) and (3) (carboxyl-ion system, B), and these are termed nonconverging series. Less symmetric chromophores including those of the dipolar amidic systems ((4)–(6), chromophore C) show markedly reduced shifts as each vinyl unit is added (converging series). In the dyes the absorption shift is related to the degree of asymmetry between the structures of the heterocyclic terminal groups. Infrared absorbing dyes are primarily derived from symmetric structures; the absorption spectra for the vinylogous series of dye (2) (Fig. 2) are typical, with broader absorptions for the dyes with $n = 4, 5$.

Table 1. Absorption of Vinylogous Dyes^a

Dye	Absorption data	<i>n</i>			
		0	1	2	3
<chem>CCN1=CC=C(C=C1)C=CC2=CC=CC=C2S2</chem> (2)^b	<chem>CCN1=CC=C(C=C1)C=CC2=CC=CC=C2S2</chem> (2)^b	423	557	650	758
$n = 0$ [2197-01-5]	λ_{max} , nm				
$n = 1$ [905-97-5]					

Fig. 2. Absorption spectra for a vinylogous thiacyanine series (dye (2), $n = 0$ to 5).

Terminal Groups. The cyanines (A), oxonols (B), and merocyanines (C) were originally considered as polymethine dyes with two heterocyclic terminal groups. Hemicyanines were defined as dyes with one heterocyclic and one noncyclic terminal group. Dye bases such as (5), Table 1, were the nonalkylated analogues of cyanine dyes like (2). Currently, dyes without heterocyclic terminal groups are designated as either cyanines or polymethine dyes. In fact, almost any atom or group of atoms can function as a terminal group for dyes if the nitrogens and oxygens in the primary chromophores (A), (B), and (C) are replaced by electronically equivalent atoms. Dyes from novel terminal groups are quite numerous (22–24). However, the fundamental concepts and perhaps the largest class of useful cyanines are derived from dyes with heterocyclic terminal groups. The heterocycles are of two principal types: basic or electron-donating, and acidic or electron-accepting. Typical examples of terminal groups are shown in Figure 3.

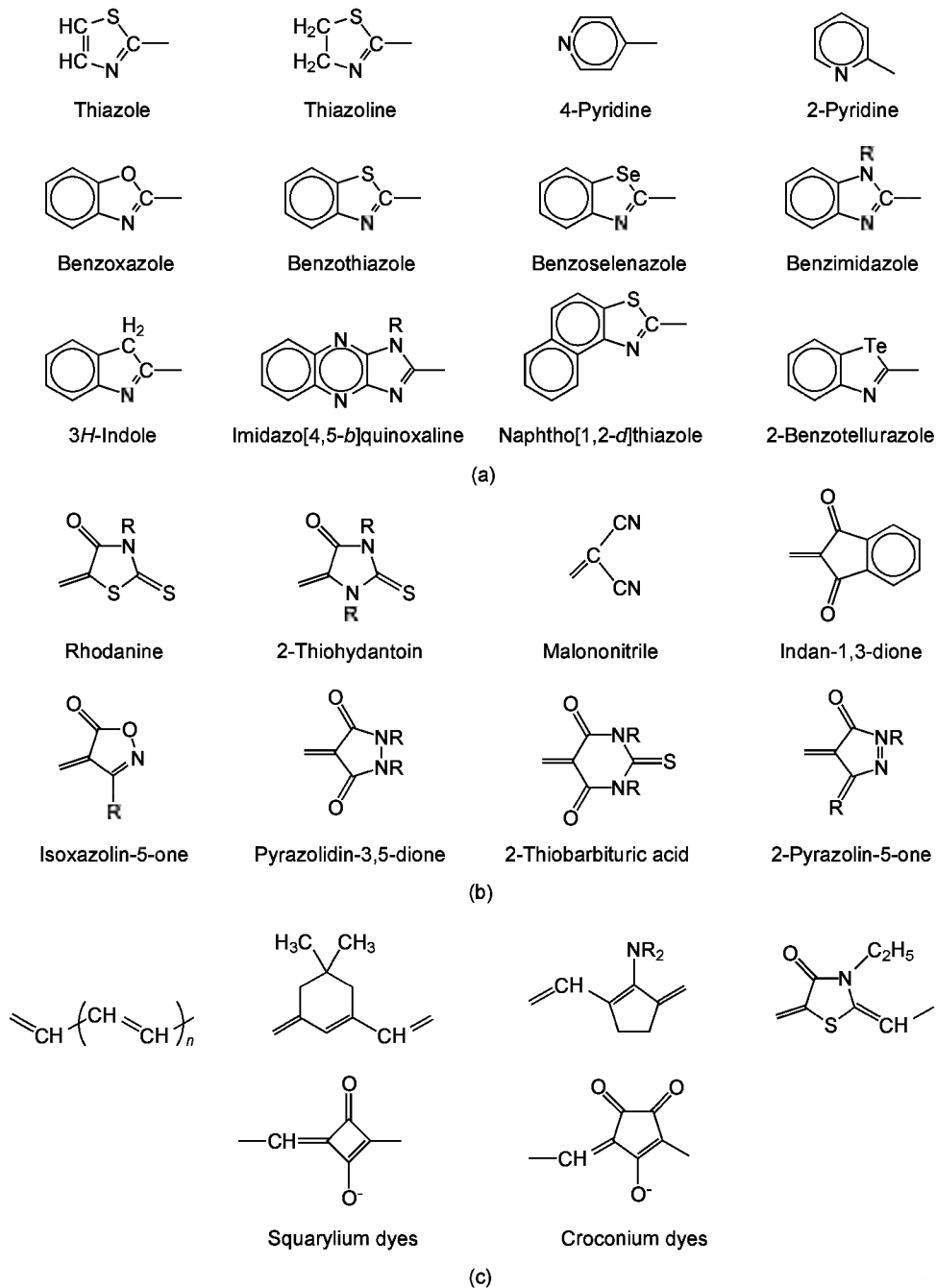
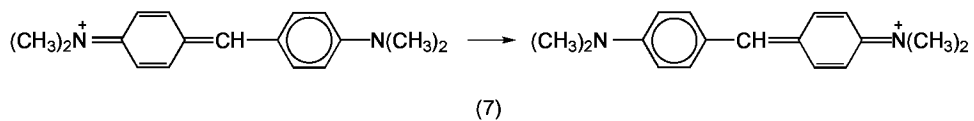
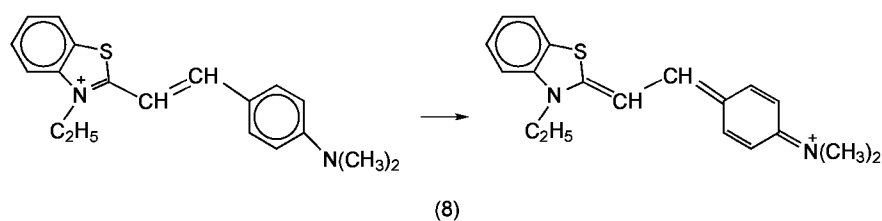


Fig. 3. Examples of nuclei occurring in important cyanine dyes. (a) basic terminal groups for cyanines and merocyanines; (b) acidic terminal groups for merocyanines and oxonols; (c) polymethine chains in longer wavelength dyes.

Basic Heterocycles. In addition to the early benzothiazole dyes, eg, (2), other heterocyclic thiazoles as well as related oxazoles, pyrroles, and imidazoles were subsequently used for cyanines. When two different terminal groups were incorporated, certain unsymmetrical dyes absorbed at unexpectedly short wavelengths, whereas the absorption of others more closely approximated the mean wavelength for the related symmetrical dyes. These observations resulted in the concept of deviation, which related the absorption characteristics of unsymmetrical dyes to the electron-donating abilities (basicities) of the various heterocycles. It was suggested that for the weakly basic terminal *p*-dimethylaminophenyl group the importance of the two extreme amidinium resonance structures would depend on the basicity of the other terminal group (2). For the symmetrical dye (7), Michler's hydrol blue [14844-71-4] the two resonance structures are equivalent, but in the unsymmetrical carbocyanine "styryl" dye [41017-17-8] (8) the structure with the charged heterocycle is favored particularly for good electron donors (highly basic heterocycles).





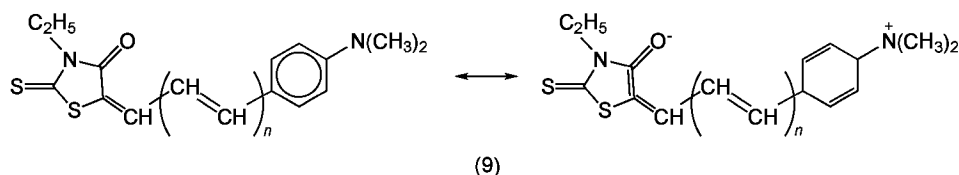
Since highly basic heterocycles accommodate the positive charge readily and maintain the aromatic nature of the dimethylamino benzene group, the resulting bond alternation induces a polyene character to the dye chromophore, and the absorption is shifted accordingly to a shorter wavelength.

A quantitative expression of these observations is shown in equation 1, where λ_{obs} is the observed absorption maximum for the unsymmetrical carbocyanine and λ_1 is the arithmetic mean (isoenergetic wavelength) for the absorption maxima of the related symmetrical dyes.

$$\text{deviation} \quad \Delta\lambda \quad \lambda_1 \quad \lambda_{\text{obs}} \quad (1)$$

Many deviations are tabulated in Ref. 16 along with the structures of the basic heterocycles. The higher the deviation, the more readily the nucleus accommodates a positive charge.

Acidic Heterocycles. A similar classification is made for the acidic electron-accepting terminal groups used in dipolar (merocyanine) chromophores. The unsymmetrical dyes again incorporate the *p*-dimethylaminophenyl group, connected to the acidic group (Fig. 3) by one or three methine carbon atoms as in the merocyanine(9), $n = 0$ [23517-90-0]; $n = 1$ [42906-02-5]; $n = 2$ [66037-49-8]; $n = 3$ [66037-48-7].



For the unsymmetrical dye (a benzylidene if $n = 0$), the nonpolar resonance form was expected to be the dominant one, except when highly electron-accepting terminal groups were present. Nonpolar dyes exhibited the characteristics of polyenes, absorbing at much shorter wavelengths than expected from the arithmetic mean absorption for the two symmetrical dyes. Because of this, the least acidic (electron-accepting) groups showed the highest deviations. The least acidic heterocycles are typified by rhodanines and hydantoin (Fig. 3); the more acidic are isoxazolones and pyrazolidinediones. Acidic terminal groups with intermediate deviations include indandiones and malononitrile.

Dyes derived from these fundamental basic and acidic terminal groups are in current use today as photographic spectral sensitizers [100471-81-6] (10), chemotherapeutic dyes [54444-00-7] (11), laser dyes [53655-17-7] (12), and biological stains [7423-31-6] (13) (Fig. 4).

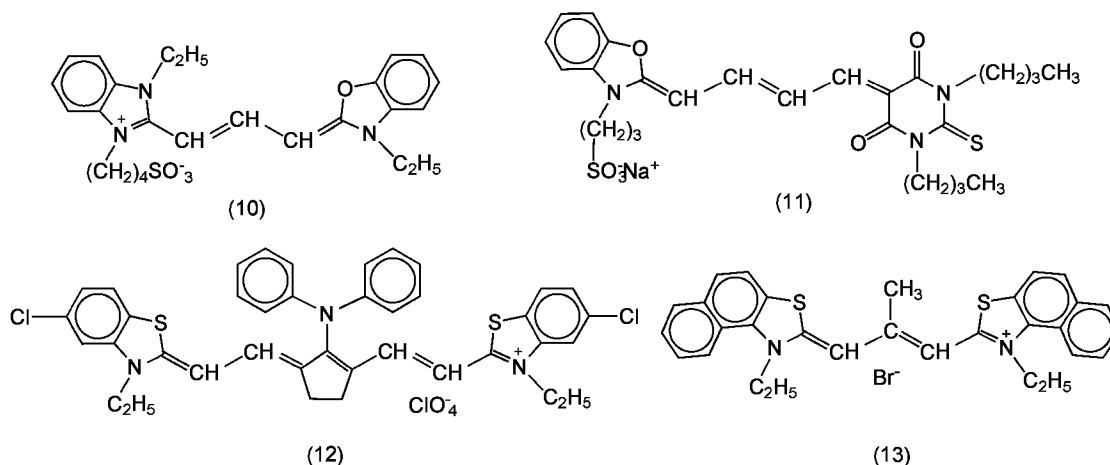


Fig. 4. Useful cyanine dyes. (10) Green spectral sensitizer; (11) Merocyanine 540; (12) IR-140 laser dye; (13) Stains-All.

Solvent Sensitivity. There is a large influence of solvents on absorption spectra for dyes, particularly for the merocyanines, (C). Merocyanines with widely different terminal groups illustrate the general pattern of solvent effects on the absorption maxima and peak intensities (24): Weakly polar dyes like (4) (Table 1) are red-shifted (longer wavelength) and show increased extinction coefficients (ϵ_{max}) as solvent polarity increases; and highly polar dyes like (6) are blue-shifted and show decreased ϵ_{max} values for increasingly polar solvents. The large effects of the environment (solvent) around a dye on the absorption spectra led to the synthesis of hundreds of dyes to clearly document relations between structure and spectra. Early interest (25) in dye (6), *p*-hydroxystyryl dyes, and other dyes (24), as solvent polarity indicators, continues, and experimental attention (26,27) has led to the derivation of linear free energy relationships for strongly conjugated systems (28).

Both merocyanine and cyanine dye classes can exhibit solvent sensitivity. For dyes with long chromophoric chains of $-\text{CH}=\text{CH}-$ (methine) groups, the conformation and thus the absorption of charged dyes may change as a function of solvent. Dye (6) with $n = 3$ (Table 1), is a solvent-sensitive, infrared dye that absorbs at longer wavelength in pyridine than in pyridine-water mixtures. The all-*trans* conformations increase in pyridine where the polar terminal groups are separated by the maximum distance. The design of infrared dyes with increased absorption at long wavelengths incorporates conformation-restricting groups in the polymethine chain (see examples, Fig. 3). Additionally, the solvent sensitive "allopolar" (partly polar) cyanine dyes have been investigated extensively (11). These dyes exist in two distinct conformations: one with complete charge separation (holopolar, *H*) and the other with at least one nonionized resonance form (meropolar, *M*). The principal absorption shifts that occur between polar and nonpolar solvents are assigned to these isomers; the holopolar form predominates in polar solvents. The charge separated (holopolar) form of an allopolar infrared dye [66037-43-2] is shown as dye (16) in Figure 5.

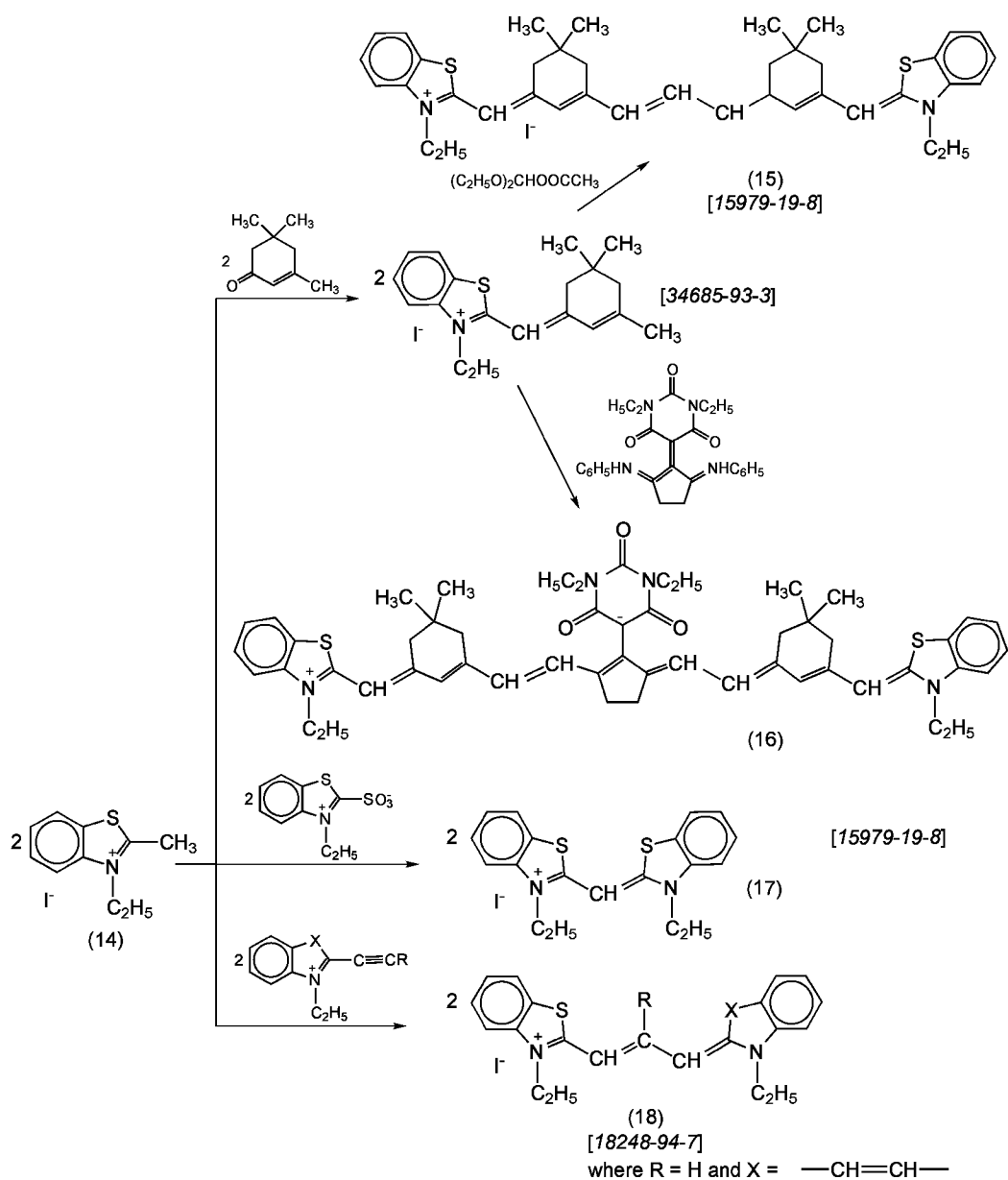


Fig. 5. Nonoxidative cyanine syntheses. Reactions of the methylene base from (14) with electrophilic reagents.

Synthesis of Cyanines and Related Dyes

General synthetic methods were developed after 1920 and extended to many new systems. Oxidative syntheses of dyes are primarily of historical interest (1), whereas nonoxidative syntheses are the most versatile and employ varied combinations of nucleophilic and electrophilic reagents. One review lists references for the synthesis of dyes prepared before 1959 (12), and another review provides supplemental references to more recent compounds (13). Many nucleophilic and electrophilic reagents used to synthesize cyanine and related dyes are tabulated in Reference 16.

Nonoxidative Syntheses. Dye syntheses are characterized as condensations, "two intermediates reacting under suitable conditions with elimination of some simple molecule" (11). Typical synthetic reactions, often combining reagents of quite different structure, are illustrated in Figures 5 and 6. Nucleophilic methylene bases are derived from the deprotonation of quaternary salts like 3-ethyl-2-methylbenzothiazolium iodide [3119-93-5] (Fig. 5, (14)). The methylene bases react with a variety of electrophilic reagents to form dyes in which the nucleophilic reagent (14) is the terminal heterocyclic group for the chromophore. The quaternary salts are also converted with ortho esters or anilides to generally useful electrophilic reagents, eg, the intermediate 3-ethyl-2-[2-(acetylphenylamino)ethenyl]benzothiazolium iodide [35080-47-8] (19) (Fig. 6).

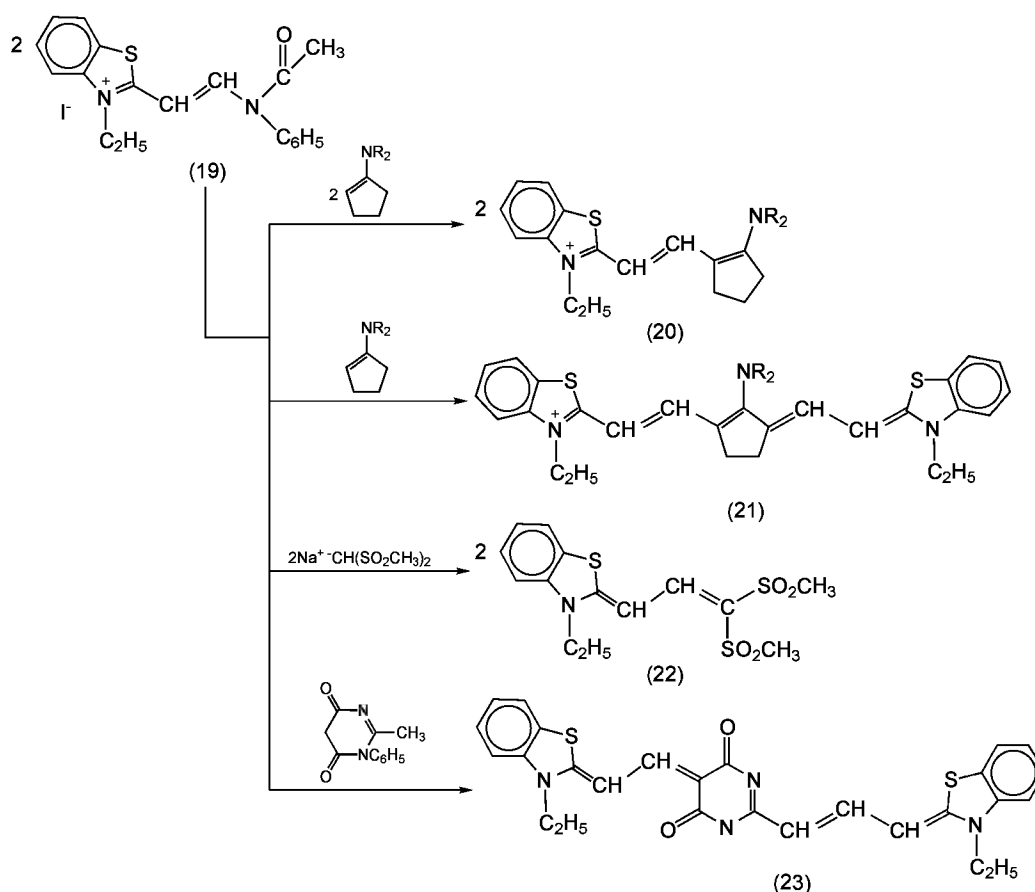
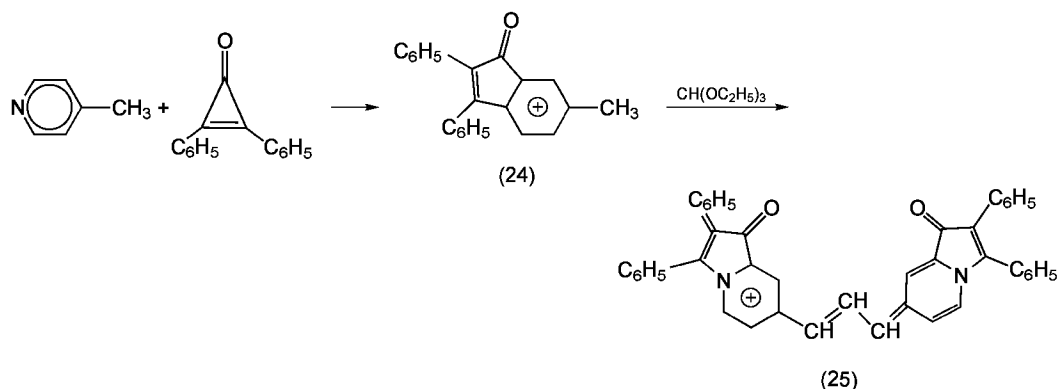
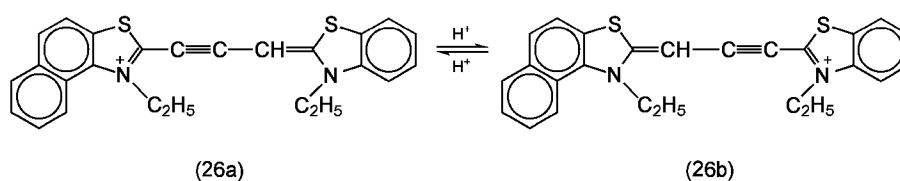


Fig. 6. Nonoxidative cyanine syntheses. Reaction of an anilide (19) with nucleophilic reagents.

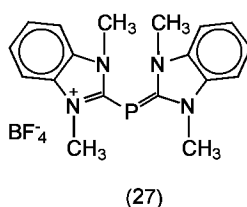
Reaction with various nucleophilic reagents provides several types of dyes. Those with simple chromophores include the hemicyanine iodide [16384-23-9] (20) in which one of the terminal nitrogens is nonheterocyclic; enamine tricyanone iodide [16384-24-0] (21) useful as a laser dye; and the merocyanine [32634-47-2] (22). More complex polynuclear dyes from reagents with more than one reactive site include the trinuclear BAB (Basic-Acidic-Basic) dye [66037-42-1] (23) containing basic-acidic-basic heterocycles. Indolizinium quaternary salts (24), derived from reaction of diphenylcyclopropenone [886-38-4] and 4-picoline [108-89-4] provide trimethine dyes such as (25), which absorb near 950 nm in the infrared (23).



Structural variations of the reagents used in these reactions have been a primary source of progress in dye synthesis. Acetylenic reagents for cyanine dye synthesis include the well-known acetylenic quaternary salts as general electrophilic reagents for the preparation of carbocyanine dyes. A number of tautomeric pairs of acetylenic dyes have been prepared and their tautomeric equilibria determined (dyes (26a), (26b)) (29).



Investigations of phosphorus reagents led to simple phosphacyanines (chromophoric methine replaced by phosphorus) as in dye (27) [55620-79-6] and phosphacarbocyanines (30).



Ring-Closure Reactions. Some interesting dyes are prepared by ring-closure reactions at or near the dye-forming step of a synthetic sequence. The structural identity of thiacyanine was originally established by the reaction of diethyl malonate and *o*-aminothiophenol.

Highly fluorescent, rigidized dyes form readily (31). Quaternary salts such as (28) (with reactive *N*-alkyl groups) are formed readily from acrolein ($R' = H$) and heterocyclic hydrobromide salts (Fig. 7) (31). Direct formation of dyes from the uncyclized quaternary salt (28) lead to the thiacyanocyanine (29). The reactive *N*-alkyl groups in this dye are ring-closed to give either the partly rigid thiacyanocyanine (30) or the completely rigid and highly fluorescent (31).

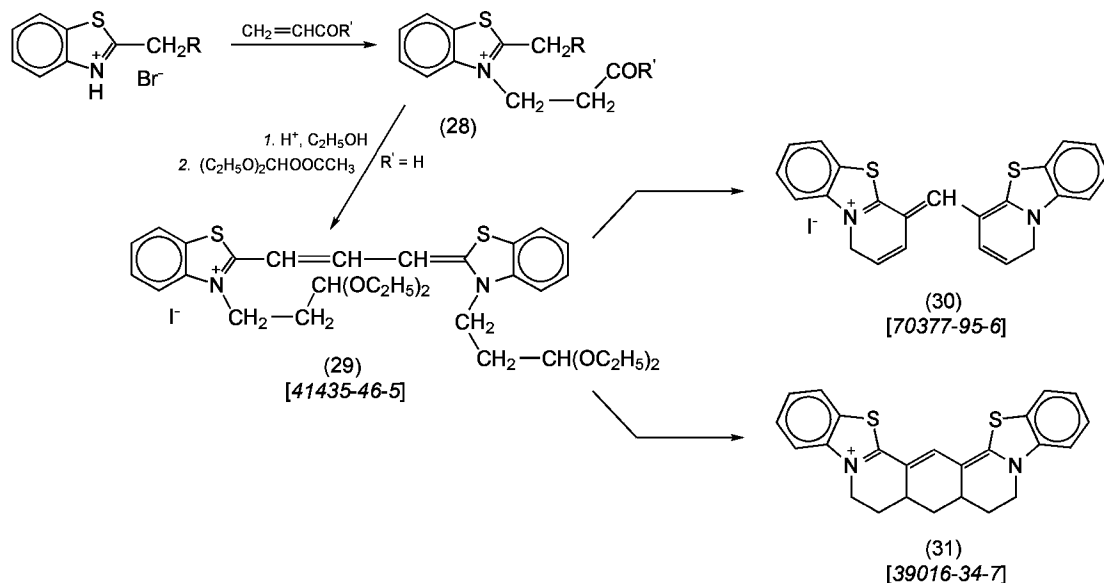


Fig. 7. Ring-closure reactions.

Physical Characterization of Ground-State Dyes

Crystal structure analyses of cyanine and related dyes are reviewed in Ref. 32. Most typical sensitizers are nearly planar, with angles of less than 15° between planes defined by heterocyclic rings. Distinct solvent of crystallization is present in most of the cationic dyes. X-ray crystal analyses also provide intermolecular data. Because of photographic use of cyanine and carbocyanine dyes, the cation-cation arrangements of most interest have been those for 1,1'-diethyl-2,2'-quinocyanine chloride [2402-42-8], 5,5',6,6'-tetrachloro-1,1',3,3'-tetraethylbenzimidazolocarboquinine iodide [3520-43-2], and 5,5'-dichloro-3,3',9-triethylthiacyanocyanine bromide [18426-56-7] (32) (see Fig. 8).

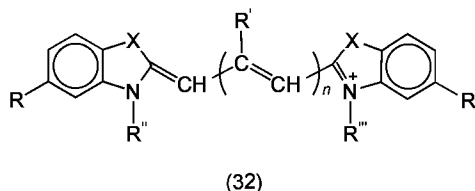


Fig. 8. Sensitizing dyes of the cyanine class. X = *N*-alkyl or chalcogens (O, S, Se, Te); R = chloro, phenyl, or additional benzene ring; R' = methyl, ethyl, or hydrogen; n = 0, 1, 2; and RPRIME, R''' = alkyl or sulfoalkyl. Solubility in methanol for a carbocyanine dye: n = 1; X = S; R = Cl; R' = ethyl. Cationic dye (R'' = ethyl; R''' = ethyl; anion = bromide) 9.5 mmol/L; neutral dye (R'' = ethyl; R''' = sulfopropyl) 3.6 mmol/L; anionic dye (R'' = R''' = sulfopropyl; cation = sodium) 25.5 mmol/L.

Oxidation-Reduction Potentials. With color-structure relationships extensively documented by synthetic efforts through 1960 (12), the electrochemical properties of dyes were selected for early correlation with the photographic effects of the cyanine dyes (33). However, a review of this early data suggested that completely reversible potentials were rare. Some apparent deviations are caused by specific substituents like nitro, but contributing factors like increased chain length for vinylogous series of dyes suggested that irreversible side reactions were rapid enough to affect the electrochemical measurements.

More recent research provides reversible oxidation-reduction potential data (17). These allow the derivation of better structure-activity relationships in both photographic sensitization and other systems where electron-transfer sensitizers are important (see DYES, SENSITIZING). Data for an extensive series of cyanine dyes are published, as obtained by second harmonic a-c voltammetry (17). A recent "quantitative structure-activity relationship" (QSAR) (34) shows that Brooker deviations for the heterocyclic nuclei (discussed above) can provide estimates of the oxidation potentials within 0.05 V. An oxidation potential plus a dye's absorption energy provide reduction potential estimates. Different regression equations were used for dyes with one-, three-, five-methine carbons in the chromophore. Also noted in Ref. 34 are previous correlations relating Brooker deviations for many heterocyclic nuclei to the pK_a (for protonation-decolorization) for carbocyanine dyes; the pK_a is thus inversely related to oxidation potential values.

Dye radicals, formed during electrochemical measurements, can be very unstable and react on very short time scales (17,35). Specifically for the dicarbocyanines (36), one-electron oxidation gives a radical dication which then dimerizes by linking the (8, 10') carbons as in Figure 9 or the (10, 10') carbons in analogous chemical steps. The 8, 10, 12 positions of the methine chain have the highest electron density in the ground state dye, and in symmetric dicarbocyanines the 8 and 12 carbons are equivalent. Thus, dimerization between electron-rich methine carbons would statistically provide 8, 8' > 8, 10' > 10, 10', whereas sterically the 10' position is the least hindered. Experimentally, the following dimers were observed for the dicarbocyanine radical dication in Figure 9: X = S, 8, 10' > 10, 10'; X = C(CH₃)₂, 10, 10'.

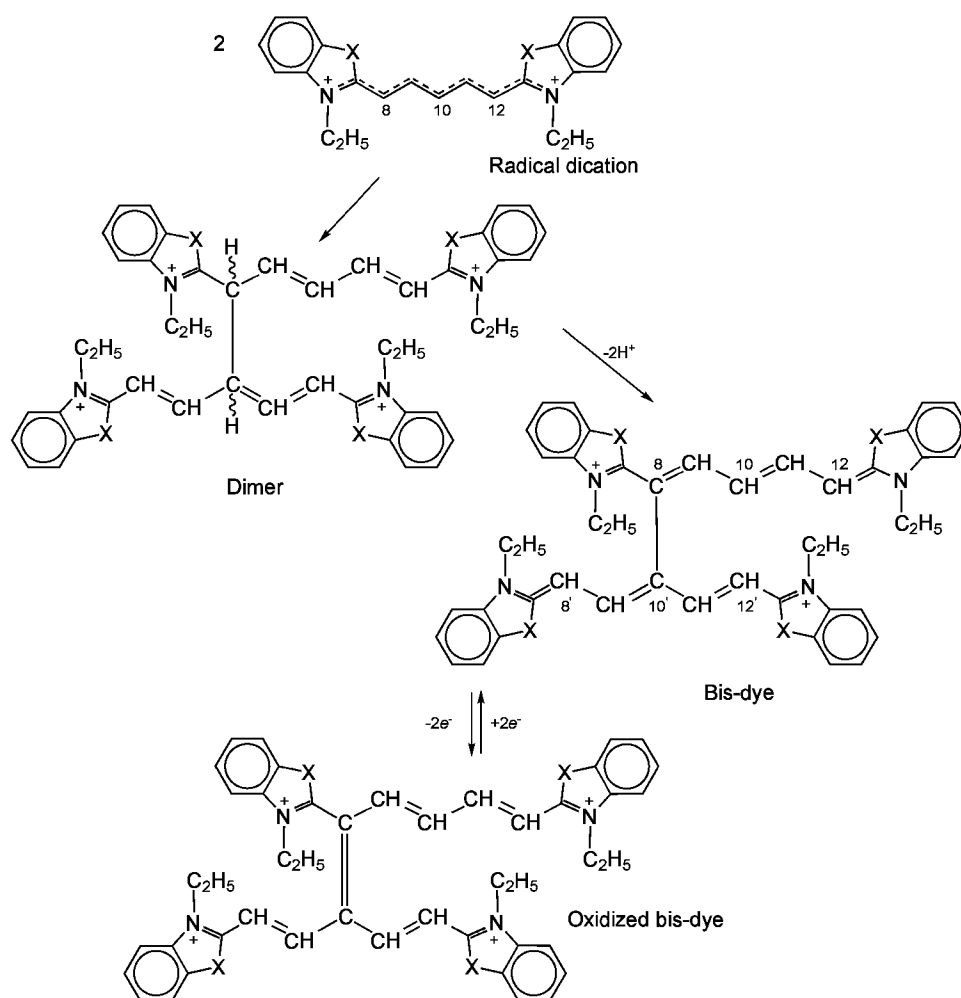
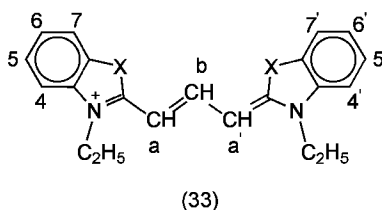


Fig. 9. Reaction pathways for oxidized dicarbocyanine dyes. Courtesy of the American Chemical Society.

Substituents on the methine chain can stabilize the dye radical cation if the substituent (like methyl) is located on the high electron density carbons. However, no significant stabilization occurs when alkyl groups are on the alternate positions (like 9, 11 for the dication in Fig. 9). Current results for several dyes including dicarbocyanines and carbocyanines indicate that electronic stabilization of the dication radical lengthens the radical lifetime and also enhances the reversibility of the dimerization process (37).

Dye Substituents. Cyanine dyes used as commercial imaging sensitizers or in Langmuir-Blodgett films are substituted to improve their sensitizing, chemical, and physical properties. In a structure such as (32) with $n = 1$, for example, the X-atom and R-substituent affect sensitization efficiency and absorption wavelength, the R'-substituent alters self-association of dye molecules, and the R'', R'''-substituents affect solubility. If R'', R''' are octadecyl, low water solubility results even for this cationic dye and good Langmuir-Blodgett films can be prepared (38,39). If R'', R''' are chosen from the short-chain alkyl or sulfoalkyl groups, the dyes with a net positive or negative charge are more soluble in protic solvents than the dye having counterbalancing positive and negative charges. In addition, the counterions of dyes with a net charge influence solubility; a cationic dye with chloride, bromide, iodide, or perchlorate anion, for example, will exhibit a wide range of solubility. Solubility data for a carbocyanine dye example are given in Figure 8.

Several types of nitrogen substituents occur in known dye structures. The most useful are the acid-substituted alkyl N-substituents such as sulfopropyl, which provide desirable solubility and adsorption characteristics for practical cyanine and merocyanine sensitizers. Patents in this area are numerous. Other types of substituents include: N-aryl groups, heterocyclic substituents, and complexes of dye bases with metal ions (iridium, platinum, zinc, copper, nickel). Heteroatom substituents directly bonded to nitrogen (N-O, N-NR₂, N-OR) provide photochemically reactive dyes.



Substituents on the methine chain are most important historically (8). These substituents cause two types of changes in the properties of the dyes. First, the absorption maxima shift depending on the inductive (electronic) effect of the substituent and its position in the methine chain. Azacyanines absorb at significantly shorter wavelength than the monomethine analogues, and related simple cyanines with phosphorus and arsenic atoms in the chain also absorb at shorter wavelengths. In azacarbocyanines the α -aza dyes absorb at shorter wavelength than the carbon analogues. Similar effects were found for carbocyanines with α -cyano, α -formyl, and α -nitro groups, whereas groups like α -alkyl, α -alkoxy, α -alkylthio, and α -fluoro generally shift the absorption to longer wavelengths. On the other hand, β -aza and β -cyano (nitro, C₄F₇, etc) carbocyanines absorb at longer wavelengths. Secondly, substituents that change the steric properties of the dye can affect both the color and the aggregation. Aggregation and substituent effects are described in a later section. Substituents that affect the color of dyes through steric hindrance include: small alkyl groups on the methine chain, which alter equilibria between cis and trans isomers of the chromophore; bulky alkyl groups on the methine chain or the heterocyclic nitrogen, which cause the chromophore to become non-planar; rigidizing substituents such as the six-membered carbocyclic rings in dyes (15) and (16) (Fig. 5), which decrease the number of cis isomers and enhance longer wavelength absorption by favoring the all trans isomer of infrared dyes.

Ring substituents for the carbocyanine chromophore (33) have been studied extensively. Up to 40 groups in the 5,5'- and 6,6'-positions are tabulated for the benzothiazole and benzimidazole dyes (13). For the thiacyanines, most substituents provided absorption maxima at longer wavelengths. Halogens, alkoxy, and alkyl groups in either the 5- or 6-position provided similar shifts. Electron-withdrawing groups and conjugated

substituents in the 6,6'-positions show greater bathochromic effects than in the 5,5'-positions. Certain unsaturated substituents provided polymerizable dyes. Hammett ζ - ρ relations for many of these ring substituents correlate with data other than absorption maxima. Both pK_a and polarographic reduction potential correlate with ζ values for several dye series, $X = N-R$, O, S, and $C(CH_3)_2$ in (33).

Additional benzenes or heterocycles have been fused to the essential chromophores of thiazolocarboyanine, imidazolocarboyanine, and other dyes. These new rings are similar to substituents, since fusing benzene rings to either the 4,5- and 4',5'-positions or the 5,6- and 5',6'-positions in (33) gives a similar color change. The resulting electrochemical properties (and, therefore, chemical reactivity) depend strongly on the position of the benzene ring. Many variations in heterocyclic structure are obtained in this way, and some examples are included in Figure 3.

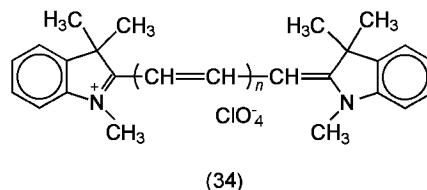
Photophysical Properties

The Electronic Structure of Sensitizers. Large conjugated molecules such as the cyanines and related dyes are well described by general quantum theories (40) as well as by resonance concepts (2,11). Both prove useful in the design and understanding of dyes. More recently, the alternating pi-electronic charge characteristic of a polymethine chain was used to define a "polymethine state", as distinct from olefinic and aromatic conjugated systems (28). Considering the complete electronic structure of dye molecules, the filled orbitals of the ground states contain sigma (σ), pi (π), and lone-pair (n) electrons. Antibonding orbitals are typically pi* (π^*) and sigma* (σ^*). The long-wavelength transitions for cyanine dyes are generally high-extinction transitions involving the π -electrons and show long-axis polarizations in cyanine dyes, as determined by dichroic absorption of polarized light (stretched polymer films with dye) and by fluorescence polarization. Shorter-wavelength transitions are polarized both parallel and perpendicular to the long axis.

Most cyanines show prominent, hypsochromic vibrational shoulders at shorter wavelength associated with the long-wavelength electronic transition (Fig. 2). These vibrational shoulders include one or two vibrational quanta ($0 \rightarrow 1$; $0 \rightarrow 2'$), in addition to the absorption energy at $\lambda_{\max}$ ($0 \rightarrow 0'$). Intensity ratios for the $0 \rightarrow 0'$ $0 \rightarrow 1'$ bands increase as the chromophoric chain is lengthened from $n = 0$ to $n = 3$ in the thiacyanine series (2) (Fig. 2). Distinct vibrational-frequency progressions are resolvable in the low-temperature absorption spectra of azacyanine-, cyanine- and carboyanine-dyes. Dyes with very long methine chains show short-wavelength absorptions that are not vibrational bands (Fig. 2), but are due to nontrans isomers of the methine chain.

Excited-state properties of the cyanine and related dyes are complex. Most cyanine dyes exhibit small Stokes shifts for fluorescence maxima. Typical carboyanines (1) with $n = 1$ show 14- to 16-nm shifts in methanol solution with low quantum efficiencies for fluorescence (Φ_F) of less than 0.05. The dicarboyanine analogues also show small Stokes shifts but higher quantum yields ($\Phi_F = 0.3$ -0.5).

Data on fluorescence, phosphorescence, excited-state lifetimes, transient absorption spectra, and dye lasers are tabulated in Ref. 16. The main nonfluorescent process in cyanine dyes is the radiationless deactivation $S_1 \rightarrow S_0$. Maximum singlet-triplet interconversion (Φ_{ST}) in methanol for carboyanines is about 3% ($\max \Phi_{ST} > 0.03$), and the sum $[\Phi_F + \Phi_{ST}]$ is less than 0.10.



Dicarboyanine and tricarboyanine laser dyes such as structure (1) ($n = 2$ and $n = 3$, $X = \text{oxygen}$) and structure (34) ($n = 3$) are photoexcited in ethanol solution to produce relatively long-lived photoisomers (10^{-4} - 10^{-3} s), and the absorption spectra are shifted to longer wavelength by several tens of nanometers (41,42). In polar media like ethanol, the excited state relaxation times for tricarboyanine (34) ($n = 3$) are independent of the anion, but in less polar solvent (dichloroethane) significant dependence on the anion occurs (43). The carboyanine from structure (34) ($n = 1$); exists as a tight ion pair with borate anions, represented $RB(C_6H_5)_3$, in benzene solution; photoexcitation of this dye-anion pair yields a new, transient species, presumably due to intra-ion pair electron transfer from the borate to yield the neutral dye radical (ie, the reduced state of the dye) (44).

Aggregation. The most notable spectral property of the cyanine dye class is the ability to self-associate or aggregate. For the more symmetrical cationic dyes, these aggregates can exhibit exceptionally narrow, high-extinction absorptions at longer wavelength than the monomeric dye absorption (Fig. 10, M — monomer). Such aggregates, termed J-aggregates for Jelley (45) who first observed them, provide exceptionally color-selective and efficient sensitization for color imaging systems. Hypsochromically shifted aggregates (dimers, trimers, H-aggregates) also occur, but their absorptions are typically as broad as the monomer (Fig. 10).

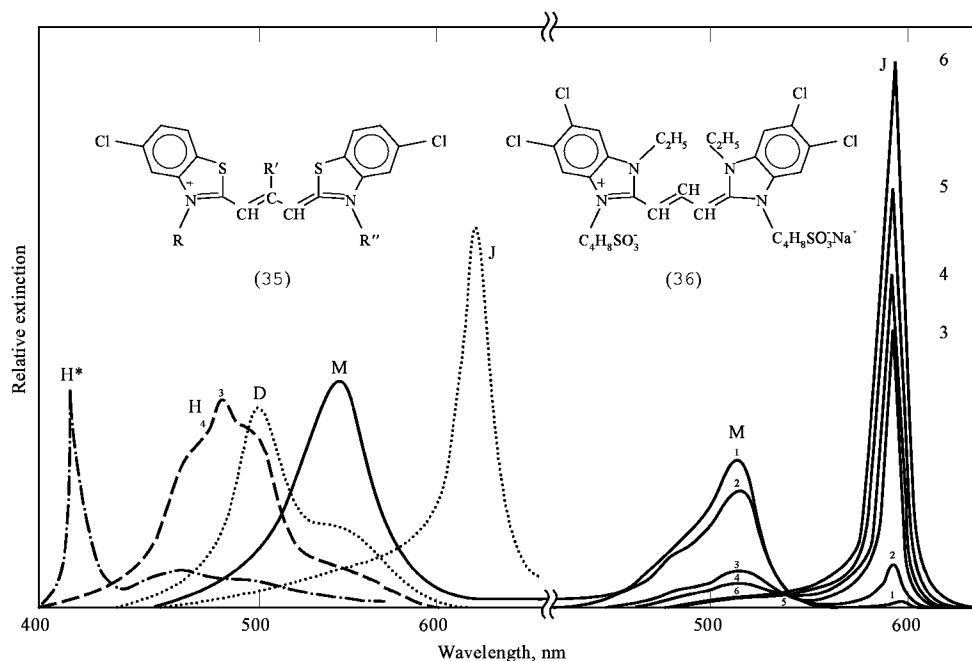
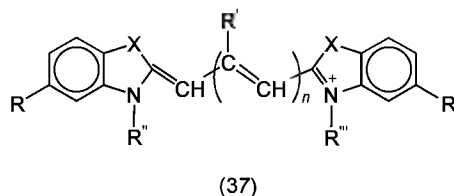


Fig. 10. Relative absorption of aggregates. For dye (35) with $R = C_2H_4COO^-$, $R' = CH_3$, $R'' = C_2H_4COOH$, 5,5'-dichloro-3,3'-di(2-carboxyethyl)-9-methyl-thiazolocarboyanine hydroxide [18281-98-6], the absorptions at shorter wavelengths than the monomer (M) band occur in various solvents where D — dimer; H — 3,4 — trimer; and H* — highly ordered hypsochromic aggregate. For the corresponding sulfonate salt, dye (35) with $R = (CH_2)_4SO_3^-$, $R' = C_2H_5$, $R'' = (CH_2)_4SO_3^- Na^+$ [42905-44-2], the J-aggregate occurs. 5,5',6,6'-Tetrachloro-1,1'-diethyl-3,3'-disulfoethylbenzimidazolocarboyanine sodium salt [18462-64-1] shows a high tendency toward J-aggregation as dye concentration increases from low (curve 1) to high (curve 6).

The π -systems of dyes can overlap strongly if the long dimensions of the chromophores are parallel and the methine chain heterocycles are stacked face-to-face (as in a deck of cards) and closely spaced. Early synthetic efforts in carbocyanine dyes focused closely on substituent patterns to enhance J-aggregation and also decrease H-aggregation (11,12). Considering generalized dye structure (37) with $n = 1$, for example, the central methine-chain substituent R' can be an ethyl or phenyl group to enhance J-aggregation for benzothiazole dyes ($X = \text{sulfur}$) such as (35). The $R' = \text{methyl}$ analogue favors dimerization and H-aggregation, as shown in Figure 10. Benzimidazole dyes [such as (36), (37), $X = \text{N-alkyl}$] are more crowded, so that J-aggregation is most favored for $R' = \text{hydrogen}$.



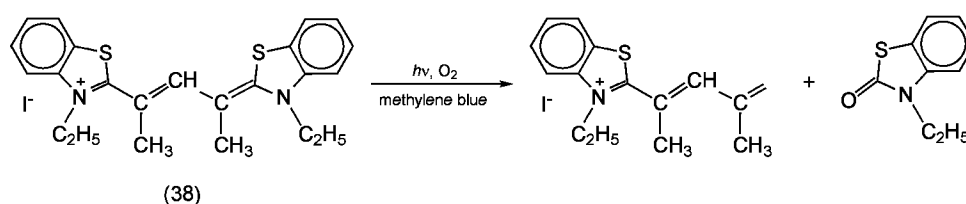
X-ray crystal structures (32,46) and molecular mechanics calculations (47,48) now provide specific data about intermolecular spacings between associated dye molecules.

In a more general description of the relation between these structures and their tendency to aggregate (11), compact dyes were defined as structures that were planar but with tightly packed substituents from a steric hindrance view-point; these dyes aggregated readily. Less hindered loose dyes (perhaps having several conformations) or more hindered crowded dyes (probably non-planar) do not aggregate as well. Of course, other groups such as the R-substituents also influence aggregation. For the benzoxazole dyes (37) ($X = \text{oxygen}$, $n = 1$), appropriate R-substituents (on each benzene ring) can enhance J-aggregation (producing spectral absorption near 550 nm or favor the monomeric dye (spectral absorption near 510 nm and sensitization in the short green spectral region).

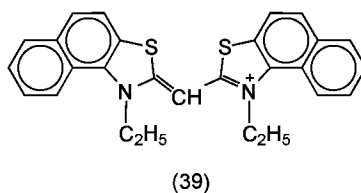
Merocyanine and hemicyanine dyes are often components of thin-film multilayers such as Langmuir-Blodgett films, where tight packing of dye molecules within a layer reinforces the self-association tendencies of the dyes. Chromophores, which orient in the films due to the method of preparation, can exhibit very strong second-order nonlinear optical properties (49–51). Symmetrical cyanine dyes in polymer films show large third-order nonlinear optical properties associated with their large conjugated π -electron systems. A tricarbocyanine (heptamethine) analogue of Williams' cyanine (Fig. 1), at 10–50% weight concentrations in an ionic polymer (sodium polystyrene sulfonate), has good processability and a high third-order nonlinear optical coefficient (52). The formation, stability, and useful properties of these dyed layers shows worldwide research interest (39,52–58).

Reactivity of Cyanine and Merocyanine Dyes

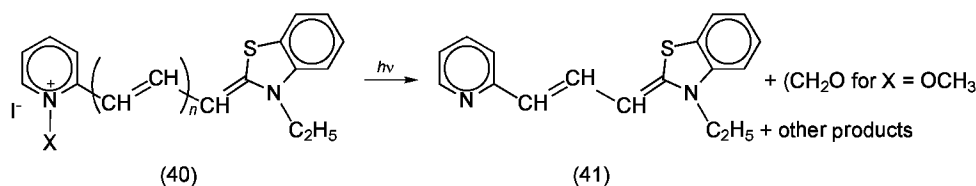
Photoreactions. Photooxidation and photoreduction of dyes have been less studied for the cyanines and merocyanines than for noncyanine dye classes. Perhaps this is due to the low singlet-triplet intersystem crossing yields and the high propensity of cyanines to undergo rapid $S_1 \rightarrow S_0$ radiationless deactivation (59). Low quantum yield photofading reactions in solutions of cyanine dyes are observed under both oxidative conditions (oxygen, reducible metal ions) and reductive conditions (ascorbic acid, gelatin). Direct and sensitized photooxidations of cyanine dyes have been investigated (60). Rate constants k_{ps} for sensitized dye bleaching are generally parallel to electrochemical oxidation potentials (ie, thiaticarbocyanine [3071-70-3] (2), $n = 3$ reacts more readily than thiaticarbocyanine). The products from these reactions, including those shown in the reaction of 3,3'-diethyl-8,8'-dimethylthiaticarbocyanine iodide [62222-86-0] (38), are consistent with 1,2-addition of singlet oxygen followed by cleavage of the $C-C$ bond between the first methine carbon and the heterocyclic ring.



3,3'-Diethyl-naphthothiacyanine iodide [18474-32-3] (39) is photodegraded more readily as the H-aggregate, $\Phi = 10^{-2}$ as compared to 10^{-6} for the monomer. Surfactants help deaggregate dyes in solution. For a thiaticarbocyanine chromophore, the deaggregation is extensive enough to produce two orders of magnitude improvement in photostability (61).



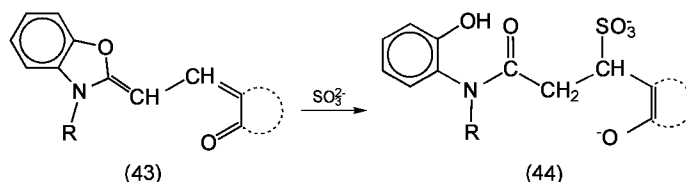
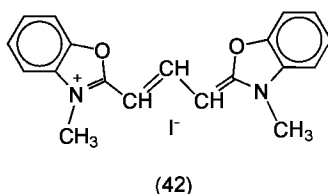
The tellurium analogues of cyanine dyes, tellurapyrylium trimethine dyes, are photocatalysts for the conversion of water to hydrogen peroxide (62). The tellurium (II)–tellurium (IV) cycle avoids the use of a sacrificial electron donor. Photoreactive substituents on quarternary nitrogen atoms allow typical synthetic reactions to be used while retaining the photosensitivity in the dye form (63). Both $N = \text{OR}$ and $N = \text{NR}_2$ for the $N-X$ part of (40) provide photobleachable and thermally bleachable dyes. The photolytic reaction for the N -methoxy dyes ($n = 1$, $X = \text{OCH}_3$, [37829-75-7]) involves initial homolytic cleavage of the $N-O$ bond and leads primarily to dye base [61109-40-8] (41), chain-substituted analogues of (41), and other heterocyclic products.



Chemical Reactions of Dyes. Decolorization is important for cyanines used in imaging materials. Understanding decolorization provides clues to dye reactions that may cause degradation of imaging materials during preparation and storage. For many dyes, protonation of the methine chain occurs readily and reversibly (64). Highly basic carbocyanine dyes like those from benzimidazole (eg, 36) protonate so readily that this provides a practical decolorization method.

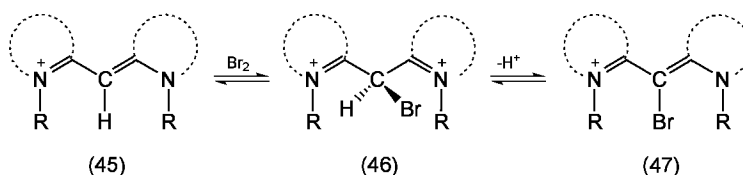
The irreversible reactions of vinylogous cyanine dyes in basic media include removal of the substituent on the heterocyclic nitrogen by amines or strong base (a general preparative method for the related dye bases (65)). Substituted *N*-alkyl groups with acidic β -protons (eg, 2-carbethoxyethyl) allow dealkylation under milder conditions through reverse Michael reactions. Other carbocyanines from pyrrole or pyrrocoline are sensitive to small amounts of base at ambient temperature. This high reactivity is desirable for decolorizing sensitizers and filter dyes during photographic processing, but stable solutions of these dyes usually require dissolution of the dyes in slightly acidified solvents (see DYES, SENSITIZING).

Benzoxazole dyes exhibit irreversible degradations that involve opening of the oxazole (66). Oxacarboyanines, eg, 3,3'-dimethyloxacarboyanine iodide [48198-86-3] (42), react most readily with aqueous acid, whereas benzoxazole merocarbocyanines (43) react with sulfite or hydroxide ion to produce ring-opened products such as (44).



The rate of reaction with sulfite is ca 100 times faster than the rate of reaction with hydroxide, and decreasing the Brooker basicity of the acidic nucleus in the merocarbocyanine increases the rate of reaction.

The reactions of halogens with cyanine dyes proceed readily at silver halide surfaces and in solution (67,68). The equilibria and halogenation products for simple cyanines from quinoline and benzothiazole have been reviewed (68). Initial halogenation is analogous to protonation and decolorizes the simple cyanines (45). Proton loss from (46) produces chain-halogenated dyes like (47) that can absorb at significantly longer wavelength than the unsubstituted cyanines. Reactions of simple cyanines, carbocyanines, and trinuclear cyanines with excess halogen have also been studied.



Innovative Uses of Dyes

Recent reviews cite both patents and scientific articles that indicate a widening spectrum of uses for the cyanine dye classes (and other dyes) in innovative technological areas (8–10,19–21,52,69–78). There are certain key elements in designing cyanines and merocyanines for the following uses: (1) long wavelength absorption accomplished with symmetrical di- and tricarbocyanine chromophores; (2) solubility modification by substituent variations without significantly affecting the other properties; (3) well-documented ways to enhance the chromophore's response to its environment (solvent, voltage, substrate) in highly unsymmetrical dyes; and (4) control through dye structure over the occurrence of aggregation on many surfaces and in appropriate solution environments.

Symmetrical, long-chain cyanine dyes for laser applications provide output from 680 to 980 nm (76). Although these dyes were obtained through early screening procedures, infrared dyes for other technologies use similar structures. A long-chain indolenine-type cyanine dye, general structure as in dye (34), has been described as the sensitizer in optical disk memories (77).

Several hundred compounds have been applied as voltage-sensitive, optical probes of neuronal activity in the brain (74). The merocyanines and styryl dyes exhibit high voltage sensitivity, but it is now realized that optimal dye design depends on the specific experimental substrate being studied, the variability among preparations of a given substrate, and in some cases the dye sample's purity. Whereas these voltage sensitive dyes are desired to have minimal damaging effects under illumination, the goal of photodynamic therapy is a cytotoxic behavior where the dyes are present. Intense absorption in the range 600–900 nm and proper solubility/adsorption properties are required, so that dyes will accumulate in solid tumors, for example, but not be retained by normal tissue or serum. Subsequent irradiation will then be selective to the dye-containing substrate. Other dyes are designed to improve detection of ribonucleic acid (RNA) when a preparation is stained with a dye (73) and to improve bioassays by reaction of isothiocyanate-substituted, near-infrared dyes with amino moieties or proteins (77).

Optical properties of cyanines can be useful for both chiral substituents/environments and also third-order nonlinear optical properties in polymer films. Methine-chain substituted dicarbocyanines have been prepared from a chiral dialdehyde: (S)-(+)-2-*sec*-butylmalonaldehyde [127473-57-8] (79), where the chiral properties are introduced via the chiral *sec*-butyl group on the central methine carbon of the pentamethine (dicarbocyanine) chromophore. For a nonchiral oxadicalcarbocyanine, the dimeric aggregate form of the dye shows circular dichroism when trapped in γ -cyclodextrin (80). Attempts to prepare polymers with carbocyanine repeat units (linked by flexible C_5H_{10} chains) gave oligomers with only two or three repeat units (81). However, these materials did form films as well as being miscible with other film-forming polymers. Degenerate fourwave mixing measurements show high optical nonlinearity for a pure oligomer film.

Suppliers. Cyanine dyes are used primarily for specialty purposes: photographic sensitizers and desensitizers, laser dyes, infrared imaging, and certain medicinal applications. Because of this, their manufacture is limited to significantly smaller quantities than for fabric dyes or other widely used coloring agents. However, the photographic, laser, and medicinal uses place high demands on the degree of purity required, and the reproducibility of synthetic methods and purification steps is very important. Suppliers of cyanine dyes include manufacturers of other specialty organic and photographic chemicals: Aldrich Chemical Company (Milwaukee, Wis.), Eastman Organic Chemicals (Rochester, N.Y.), Japanese Institute for Photosensitizing Dyes (Okayama, Japan), Molecular Probes (Eugene, Oreg.), NK Dyes (Japan), Pfaltz and Bauer (Stamford, Conn.), and Riedel deHaen (Karlsruhe, Germany). More importantly, these firms provide sources of generally useful reagents which, in two or three synthetic steps (12), lead to many of the commonly used cyanine dyes.

Health and Safety Information

A wide variety of structures exist in the cyanine, merocyanine, and oxonol classes of dyes. Properties that may affect toxicity vary widely also. These include solubility, propensity to be oxidized or reduced, aggregation tendency, and diffusion through membranes. Specific acute toxicity data are listed in Table 2, and the LD₅₀ data vary widely with the test used.

Table 2. Acute Toxicity Data for Cyanine Dyes ^a

Dye name	CAS Registry Number	Oral LD ₅₀		Dermal LD ₅₀		Intraperitoneal LD ₅₀	
		Mouse, mg kg	Rat, mg kg	Mouse, mg kg	Guinea pig, mg kg	Mouse, mg kg	Rat, mg kg
indocyanine green	[3599-32-4]					60 ^{b,c}	
cryptocyanine	[4727-50-8]	25			>50	1 ^b	
DODC iodide	[14806-50-9]	25			>1000	5–10	
DTTC iodide	[3071-70-3]	400–800		>100	>1000	1	
HIDC iodide	[36536-22-8]		178(M) <100(F)		>1000		
HPTS	[6358-69-6]	2200	800–1600		1000		
IR-125	[3599-32-4]	>100					700
Q-switch 1 ^d	[21016-19-3]				>1000	1600–3200	

^a Interchemical comparisons of acute toxicity (LD₅₀) are valid only for the same species and route of administration. Except where noted in the table, all data are previously unpublished, Health and Environment Laboratories, Eastman Kodak Company, Rochester, N.Y.

^b Material Safety Data Sheet, Sigma-Aldrich Corporation, Milwaukee, Wis.

^c This LD₅₀ was derived from an intravenous injection and not an intraperitoneal injection.

^d For neodymium laser.

More general studies on toxicity were the focus of early efforts by the Ecological and Toxicological Association of the Dyestuffs and Organic Pigments Manufacturing Industry (82–83) and more recent reports (84–85). The cyanine classes of dyes did not figure prominently in these studies. Studies of cyanine dyes in biological systems have included both inhibition of cell division and growth by an oxidation–reduction series of cyanine dyes (86), and interaction of certain cationic dyes with the respiratory chain of rat liver mitochondria (87).

Powerful solvents such as dimethyl sulfoxide (common laser dye solvent) and solubilizing substituents (R'' and R''' = sulfoalkyl in structure 32) may enhance the transport of dyes in solution through skin and other membranes. Reference 88 (on laser dye solutions and toxicity) is recommended to researchers working with dye solutions. Other dyes, such as Indocyanine Green, attain useful properties (blood tracer dye) as a result of having solubilizing substituents in their structure.

BIBLIOGRAPHY

"Cyanine Dyes" in *ECT* 1st ed., Vol. 4, pp. 742–754, by G. H. Keyes and E. J. Van Lare, Eastman Kodak Company; in *ECT* 2nd ed., Vol. 6, pp. 605–624, by E. J. Van Lare, Eastman Kodak Company; in *ECT* 3rd ed., Vol. 7, pp. 335–356, by David M. Sturmer, Eastman Kodak Company.

1. C. H. Greville Williams, *Trans. R. Soc. Edinburgh* **21**, 377 (1856).
2. L. G. S. Brooker, in C. E. K. Mees, ed., *The Theory of the Photographic Process*, 1st ed., Macmillan, New York, 1942, p. 987.
3. G. Ficken, *Chem. Ind.*, 672 (1989).
4. D. M. Sturmer and A. P. Marchetti, in J. M. Sturge, V. Walworth, and A. Shepp, eds., *Imaging Processes and Materials*, Neblette's 8th ed., Van Nostrand Reinhold, New York, 1989, p. 96.
5. **U.S. Pat. 2,925,417** (Feb. 16, 1960), E. F. Elslager and D. F. Worth (to Parke-Davis and Co.); **U.S. Pat. 2,515,912** (July 18, 1950), E. Van Lare and L. G. S. Brooker (to Eastman Kodak Co.).
6. **U.S. Pat. 2,893,914** (July 7, 1959), M. C. McCowen and P. F. Wiley (to Eli Lilly and Co.).
7. **U.S. Pat. 2,895,955** (July 21, 1959), D. W. Heseltine and L. G. S. Brooker (to Eastman Kodak Co.).
8. M. Lu, N. C. Seeman, and N. R. Kallenbach, *Biochemistry* **29**, 3407 (1990).
9. J. Davila, A. Harriman, and K. S. Gulliya, *Photochem. Photobiol.* **53**, 1 (1991).
10. F. J. Duarte and L. W. Hillman, eds., *Dye Laser Principles with Applications*, Academic Press, New York, 1990, p. 433.
11. L. G. S. Brooker, in C. E. K. Mees and T. H. James, eds., *The Theory of the Photographic Process*, 3rd ed., Macmillan, New York, 1966, p. 198.
12. F. M. Hamer, in A. Weissberger, ed., *The Chemistry of Heterocyclic Compounds*, Vol. 18, Wiley-Interscience, New York, 1964.
13. G. E. Ficken, in K. Venkataraman, ed., *The Chemistry of Synthetic Dyes*, Vol. 4, Academic Press, New York, 1971, p. 211.
14. A. I. Kiptianov, *Usp. Khim.* **29**, 1336 (1960); **35**, 361 (1966); **40**, 594 (1971).
15. D. M. Sturmer and D. W. Heseltine, in T. H. James, ed., *The Theory of the Photographic Process*, 4th ed., Macmillan, New York, 1977, p. 194.
16. D. M. Sturmer, in E. C. Taylor and A. Weissberger, eds., *The Chemistry of Heterocyclic Compounds*, Vol. 30, Wiley-Interscience, New York, 1977, p. 441.
17. J. R. Lenhard, *J. Imaging Sci* **30**, 27 (1986).
18. H. Zollinger, *Color Chemistry*, VCH, Weinheim, 1987, Chaps. 3 and 14.
19. N. Tyutyulkov, J. Fabian, A. Mehlhorn, F. Dietz, and A. Tadjer, *Polymethine Dyes: Structure and Properties*, St. Klimont Ohridski University Press, Sofia, Bulgaria, 1991.
20. M. Okawara, T. Kitao, T. Hirashima, and M. Matsuoka, *Organic Colorants; A Handbook of Selected Dyes for Electro-Optical Applications*, Elsevier, Amsterdam and New York, 1988.
21. M. Matsuoka, ed., *Infrared Absorbing Dyes*, Plenum Press, New York, 1990.
22. W. Jung, ed., *Optische Anregung Organische Systeme: Internationale Farbensymposium, Elan, 1964*, Verlag Chem., Weinheim, 1966.
23. C. H. Weidner, Eastman Kodak Co., personal communication, 1992.
24. L. G. S. Brooker and co-workers, *J. Am. Chem. Soc.* **87**, 2443 (1965); previous papers in this series.
25. L. G. S. Brooker, G. H. Keyes, and D. W. Heseltine, *J. Am. Chem. Soc.* **73**, 5350 (1951).
26. S. Arai, K. Nagakura, M. Ishikawa, and M. Hida, *J. Chem. Soc. Perkin Trans. 1*, 1915 (1990).
27. M. Niedbalska and I. Gruda, *Can. J. Chem.* **68**, 691 (1990).
28. S. Daehne and K. Hoffmann, in R. W. Taft, ed., *Progress in Physical Organic Chemistry*, Vol. 18, Wiley-Interscience, New York, 1990, p. 1.
29. J. D. Mee, *J. Am. Chem. Soc.* **96**, 4712 (1974); *J. Org. Chem.* **42**, 1035, 1041 (1977).
30. K. Dimroth, *Fortschr. Chem. Forsch.*, **38**, 5 (1973); N. Gamon and C. Reichardt, *Liebigs Ann. Chem.*, 2072 (1980).
31. L. L. Lincoln and co-workers, Eastman Kodak Co., personal communication, 1974; *Chem. Commun.*, 648 (1974).
32. D. L. Smith, *Photogr. Sci. Eng.* **18**, 309 (1974).
33. S. E. Sheppard, R. H. Lambert, and R. D. Walker, *J. Phys. Chem.* **50**, 210 (1946).
34. S. G. Link, *Abstracts of the International East-West Symposium II: Factors Influencing the Efficiency of Photographic Imaging*, Oct. 30–Nov. 4, 1988, SPSE-The Society for Imaging Science and Technology, Springfield, Va., 1988, p. F-73.
35. S. Nomura and S. Okazaki, *Chem. Letters*, 2331 (1990).

36. J. R. Lenhard and R. L. Parton, *J. Am. Chem. Soc.* **109**, 5808 (1987).

37. R. L. Parton and J. R. Lenhard, *J. Org. Chem.* **55**, 49 (1990).

38. T. L. Penner, Eastman Kodak Co., personal communication, 1978.

39. P. Stroeve, D. D. Saperstein, and J. F. Rabolt, *J. Chem. Phys.* **92**, 6958 (1990).

40. S. F. Mason, in K. Venkataraman, ed., *The Chemistry of Synthetic Dyes*, Vol. 3, Academic Press, New York, 1970, p. 169.

41. C. Rulliere, *Chem. Phys. Letters* **43**, 303 (1976).

42. X. R. Zhu and J. M. Harris, *Chem. Phys.* **142**, 301 (1990).

43. M. I. Demchuk, A. A. Ischenko, Zh. A. Krasnaya, and V. P. Mikhilov, *Chem. Phys. Lett.* **167**, 170 (1990).

44. G. B. Schuster, *Pure Appl. Chem.* **62**, 1565 (1990).

45. E. E. Jelley, *Nature* **138**, 1009 (1936).

46. K. Nakatsu, H. Yoshioka, and S. Nishigaki, *Kwansei University Annual Studies (Nishinomiga, Japan)* **29**, 213 (1980).

47. R. L. Parton, H. J. Price, and I. A. Watson, *Proceedings of the 44th Annual Conference of the Society for Imaging Science and Technology, May, 1991*, IS&T, Springfield, Va., 1991.

48. M. Ohashi, K. Kasatani, H. Shinohara, and H. Sato, *J. Am. Chem. Soc.* **112**, 5824 (1990).

49. P. N. Prasad and D. J. Williams, *Introduction to Nonlinear Optical Effects in Molecules and Polymers*, Wiley-Interscience, New York, 1991, pp. 132, 222.

50. A. Ulman, *An Introduction to Ultrathin Organic Films from Langmuir-Blodgett to Self-Assembly*, Academic Press, New York, 1991.

51. T. L. Penner, Eastman Kodak Co., personal communication, 1992.

52. S. Matsumoto, K. Kubodera, T. Kurihara, and T. Kaino, *Opt. Commun.* **76**, 147 (1990); S. R. Marder, J. E. Sohn, and G. D. Stucky, *ACS Symposium Series 455: Materials for Nonlinear Optics: Chemical Perspectives*, American Chemical Society, Washington, D.C., 1991, pp. 712–714.

53. T. Kawaguchi and K. Iwata, *Thin Solid Films* **191**, 173 (1990).

54. S. Kuroda, K. Ikegami, and M. Sugi, *Mol. Cryst. Liq. Cryst.* **190**, 111 (1990).

55. J. Bernard and M. Orrit, *J. Luminesc.* **45**, 70 (1990).

56. H. Nakahara, H. Uchimi, K. Fukuda, N. Tamai, and I. Yamazako, *Mol. Cryst. Liq. Cryst.* **183**, 345 (1990).

57. W. Leupacher, A. Penzkofer, B. Runde, and K. H. Drexhage, *Appl. Phys. B.* **44**, 133 (1987).

58. F. J. Schmitt and W. Knoll, *Chem. Phys. Lett.* **165**, 54 (1990).

59. D. F. O'Brien, T. M. Kelly, and L. F. Costa, *Photogr. Sci. Eng.* **18**, 76 (1974).

60. G. W. Byers, S. Gross, and P. M. Henrichs, *Photochem. Photobiol.* **23**, 37 (1976).

61. E. Barni, P. Savarino, and G. Viscardi, *Acc. Chem. Res.* **24**, 98 (1991).

62. M. R. Detty and S. L. Gibson, *J. Am. Chem. Soc.* **112**, 4086 (1990).

63. J. D. Mee, D. W. Heseltine, and E. C. Taylor, *J. Am. Chem. Soc.* **92**, 5814 (1970).

64. A. H. Herz, *Photogr. Sci. Eng.* **18**, 207 (1974).

65. Ref. 12, p. 362.

66. F. J. Sauter, Eastman Kodak Co., personal communication, 1992.

67. A. van Beek and co-workers, *Rec. Trav. Chim.* **94**, 31 (1975).

68. R. Steiger, J. F. Reber, and H. Loeliger, *J. Imaging Sci.* **30**, 47 (1986).

69. D. Gregory, *Chem. Ind.*, 679 (1989).

70. G. Jori, *Photochem. Photobiol.* **52**, 439 (1990).

71. M. R. Detty and co-workers, *Photochem. Photobiol.*, 1991, in press.

72. E. Buncel and S. Rajagopal, *Acc. Chem. Res.* **23**, 226 (1990).

73. **U.S. Pat. 4,883,867** (Nov. 28, 1989), L. G. Lee and C-H. Chen (to Becton, Dickenson and Co.).

74. A. Grinvald, R. D. Frostig, E. Lieke, and R. Hildesheim, *Physiol. Rev.* **68**, 1285 (1988).

75. A. Grinvald, in W. M. Cowan, ed., *Ann. Rev. Neurosci.*, Vol. 8, Annual Reviews, Inc., Palo Alto, Calif., 1985, p. 263.

76. Catalog JJ-169, Kodak Optical Products, Eastman Fine Chemicals, Eastman Kodak Co., Rochester, N.Y. 14650.

77. F. Matsui in Ref. 21, p. 128.

78. M. Lipowska, G. Patonay, and L. Streckowski, *Abstracts for American Chemical Society Meeting*, Apr. 1991, Atlanta, Ga.

79. C. Reichardt and U. Budnik, *Chem. Ber.* **123**, 2023 (1990).

80. V. Buss, *Angew. Chem. I. E.* **30**, 869 (1991).

81. L. Yu, M. Chen, and L. R. Dalton, *Polymer* **32**, 1369 (1991).

82. R. Anliker, *J. Soc. Dyers Colour.* **95**, 317–326 (1979).

83. R. Anliker, in M. Richardson, ed., *Toxic Hazard Assessment of Chemicals*, Royal Society of Chemistry, London, 1986, pp. 166–187.

84. K. Hunger and R. Jung, *Chimia* **45**, 297–300 (1991).

85. S. K. Khanna and M. Das, *J. Sci. Ind. Res.* **50**, 965–974 (1991).

86. S. Zigman and P. Gilman, *Science* **208**, 188–191 (1980).

87. T. E. Conover and R. F. Schneider, *J. Biol. Chem.* **256**, 402–406 (1981).

88. H. A. Kues and G. A. Lutty, *Laser Focus, Technology Update 1*, (May, 1975).

David M. Sturmer
Eastman Kodak Company

CYANOCARBONS

Cyanocarbons are compounds having such a large number of cyano groups that the chemical reactions of the class are essentially new in kind and not shared by analogous compounds free of such groups. The unique reactivity of cyanocarbons was first recognized with the synthesis of tetracyanoethylene [670-54-2] (TCNE) (1,2). Previously, dicyanoacetylene [1071-98-3] (3) and dicyanodiacetylene [16419-78-6] (4) were the only known "per" cyanocarbons, and cyanoform [454-50-2] (5), CH(CN)₃, was the best known compound whose properties were exceptional as the consequence of a large number of cyano groups.

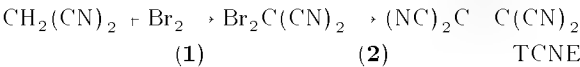
The cyano group is a powerful electron-withdrawing group. It is sufficiently small to present no great steric problem. Tetracyanoethylene must be regarded as a highly electron-deficient olefin that should be a strongly electrophilic reagent. This characteristic is reflected in the ease with which TCNE is attacked by electron-rich olefins, eg, in Diels-Alder additions and cyclobutane formation, and other nucleophiles, eg, methanol and dimethylaniline. Moreover, the affinity of TCNE for electrons is so great that the stable anion radical, TCNE^{•-}, is formed by treatment with many mild reducing agents, such as I⁻. For similar reasons, a compound with a hydrogen atom on a saturated carbon atom bearing cyanocarbon groups is highly ionized and is a very strong acid, eg, cyanoform and 1,1,2,3,3-pentacyanopropene [45078-17-9].

The most important members of the cyanocarbon class are the alkenes: tetracyanoethylene, hexacyanobutadiene, and tetracyanoquinodimethan; the alkanes: tetracyanomethane and hexacyanoethane; dicyanoacetylene; hexacyanobenzene; tetracyanoquinone; cyanocarbon acids; oxacyanocarbons; thiacyanocarbons; and azacyanocarbons. Tetracyanoethylene is described first because its chemical versatility makes it a rich source of other polycyano compounds. Moreover, an understanding of its chemistry is helpful in understanding the chemistry of other cyanocarbons.

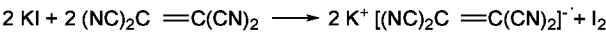
Tetracyanoethylene

Tetracyanoethylene, ethenetetracarbonitrile (TCNE) (2) has a high positive heat of formation (Table 1), but is stable to shock. It has good thermal stability and can thus be sublimed unchanged through a tube at 600°C. Tetracyanoethylene resists oxidation, but once ignited in oxygen it burns with an intensely hot flame that may be above 4000 K and is, at any rate, hotter than the oxygen-acetylene flame.

Tetracyanoethylene is colorless but forms intensely colored complexes with olefins or aromatic hydrocarbons, eg, benzene solutions are yellow, xylene solutions are orange, and mesitylene solutions are red. The colors arise from complexes of a Lewis acid-base type, with partial transfer of a π-electron from the aromatic hydrocarbon to TCNE (8). TCNE is conveniently prepared in the laboratory from malononitrile [109-77-3] (1) by debromination of dibromomalononitrile [1855-23-0] (2) with copper powder (9). The debromination can also be done by pyrolysis at ca 500°C (10).



With substances that give up an electron more readily than aromatic hydrocarbons, such as potassium, nickel carbonyl, cyanide ion, or iodide ion, complete transfer of an electron occurs and the TCNE anion radical is formed (11). Potassium iodide is a particularly useful reagent for this purpose, and merely dissolving potassium iodide in an acetonitrile solution of TCNE causes the potassium salt of the anion radical to precipitate as bronze-colored crystals.



The reaction of bis(benzene)vanadium [12129-72-5] with TCNE affords an insoluble amorphous black solid that exhibits field-dependent magnetization and hysteresis at room temperature, an organic-based magnet (12). The anion radical is quite stable in the solid state. It is paramagnetic, and its intense electron paramagnetic resonance (epr) spectrum has nine principal lines with the intensity ratios expected for four equivalent ¹⁴N nuclei (13) and may be used as an internal reference in epr work (see MAGNETIC SPIN RESONANCE).

Table 1. Physical Properties of Tetracyanoethylene^a

Property	Value
mol wt	128.1
mp, °C	200–202
bp, °C	223
density ^b , g/mL	1.318
λ _{max} , nm(ε) ^c	267 (13,600)
	277 (12,050)
heat of combustion, kJ/mol ^d	3022
heat of formation, kJ/mol ^d	628

^a Ref. 6.

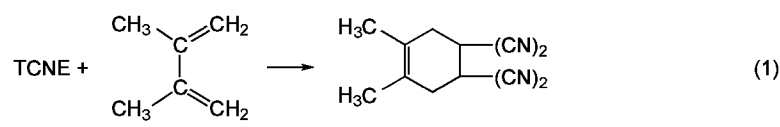
^b Ref. 7.

^c In methylene chloride.

^d To convert kJ/mol to kcal/mol, divide by 4.184.

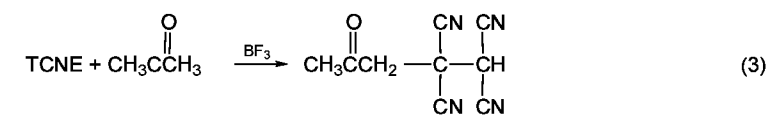
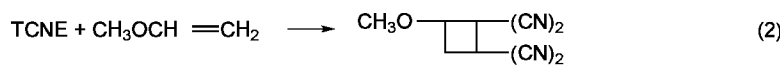
Tetracyanoethylene undergoes two principal types of reactions, addition to the double bond and replacement of a cyano group. Addition of hydrogen catalyzed by Pd gives 1,1,2,2-tetracyanoethane [14778-29-1] (14).

The Diels-Alder addition to 1,3-dienes is particularly interesting because of the exceptional ease with which it takes place, eg, see equation 1 in which 1,2-dimethyl-4,4,5,5-tetracyanocyclohexene [69155-29-9] is formed.

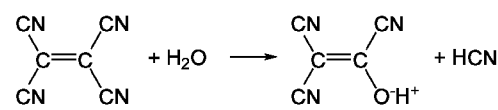


In a study of the relative rates of addition of 20 dienophiles to cyclopentadiene, TCNE was at the head of the list, eg, it added 7700 times as rapidly as maleic anhydride (15). Reaction with most 1,3-dienes takes place rapidly and in high yield at room temperature. TCNE has often been used to characterize 1,3-dienes, including those that are unstable and difficult to isolate (16).

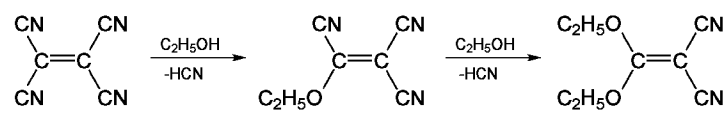
Electron-rich alkenes (eq. 2) (17), and ketones (eq. 3) (14) add readily to TCNE, as do nucleophilic radicals (14). The product in equation 2 is 1,1,2,2,-tetracyano-3-methoxycyclobutane [69155-30-2]. Addition of acetone yields 4,4,5,5-tetracyano-2-pentanone [69155-31-3].



Although a C–CN bond is normally strong, one or two cyano groups in TCNE can be replaced easily, about as easily as the one in an acyl cyanide. The replacing group can be hydroxyl, alkoxyl, amino, or a nucleophilic aryl group. Thus hydrolysis of TCNE under neutral or mildly acidic conditions leads to tricyanoethenol [27062-39-1], a strong acid isolated only in the form of salts (18).

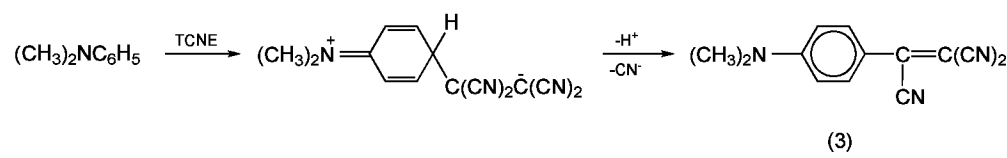


Heating TCNE with an alcohol in the presence of a mild base such as urea causes replacement of either one (19) or two (20) cyano groups by alkoxy. The products with ethanol are 1-ethoxy-1,2,2-tricyanoethylene [69155-32-4] and 1,1-bisethoxy-2,2-dicyanoethylene [17618-65-4].

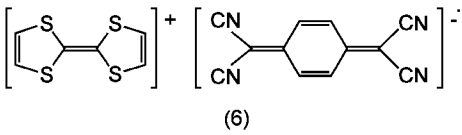


Aromatic compounds that are sufficiently nucleophilic to condense with benzenediazonium chloride and form azo compounds generally condense with TCNE, eg, the reaction of *N,N*-dimethylaniline proceeds stepwise (21,22).

The *p*-tricyanovinylarylamines such as *p*-tricyanovinyl-*N,N*-dimethylaniline [6673-15-0] (3) are highly colored substances that give bright orange, red, or blue dyeings on poly(ethylene terephthalate) and other hydrophobic fibers.

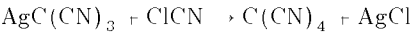


highly conducting organic compound known (10^4 S cm at 70 K) (38,39) (see SEMICONDUCTORS,ORGANIC). Moreover, the conduction is metalliclike in that it increases as temperature decreases. Superconduction properties for highly purified crystals of TTF-TCNQ at 60 K are reported (40,41) (see SUPERCONDUCTING MATERIALS).

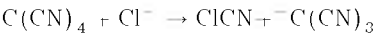


Tetracyanomethane

Tetracyanomethane [24331-09-7], methanetetracarbonitrile, is prepared by heating silver tricyanomethanide [36603-81-3] in liquid cyanogen chloride (42).

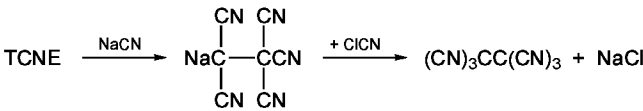


It is a very strong cyanating agent; eg, it converts chloride ion to cyanogen chloride [506-77-4].



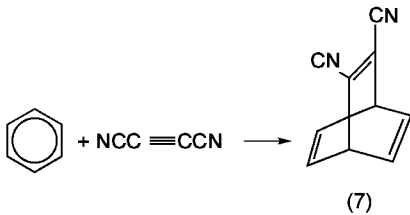
Hexacyanoethane

Hexacyanoethane [4383-67-9], ethanehexacarbonitrile, is quite unstable and readily decomposes to TCNE and cyanogen [460-19-5] (NC-CN) (43). It is prepared as follows:



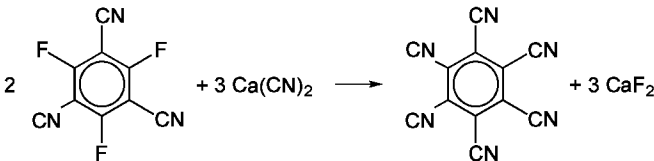
Dicyanoacetylene

Dicyanoacetylene, 2-butyndinitrile, is obtained from dimethyl acetylenedicarboxylate by ammonolysis to the diamide, which is dehydrated with phosphorus pentoxide (44). It burns in oxygen to give a flame with a temperature of 5260 K, the hottest flame temperature known (45). Alcohols and amines add readily to its acetylenic bond (46). It is a powerful dienophile in the Diels-Alder reaction; it adds to many dienes at room temperature, and at 180°C actually adds 1,4- to benzene to give the bicyclo adduct (7) [18341-68-9], $C_{10}H N_2$ (47).



Hexacyanobenzene

Hexacyanobenzene [1217-44-33], benzenehexacarbonitrile, is prepared from 2,4,6-trifluorobenzene-1,3,5-tricarbonitrile [3638-97-9] by substitution with calcium cyanide (48,49). It forms colored π -complexes with aromatic hydrocarbons.



Tetracyanobenzoquinone

Tetracyanobenzoquinone [4032-03-5], 3,6-dioxo-1,4-cyclohexadiene-1,2,4,5-tetracarbonitrile, is a remarkably strong oxidizing agent for a quinone; it abstracts hydrogen from tetralin or ethanol even at room temperature (50). It is a stronger π -acid than TCNE because it forms more deeply colored π -complexes with aromatic hydrocarbons.

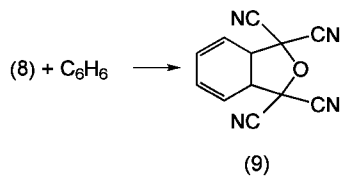
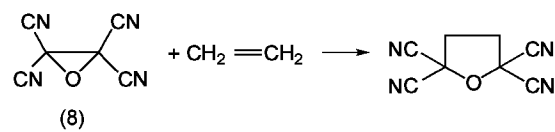
Cyanocarbon Acids

These acids (51) are organic molecules that contain a plurality of cyano groups and are readily ionized to hydrogen ions and resonance-stabilized anions. Typical cyanocarbon acids are cyanoform, methanetricarbonitrile (5); 1,1,3,3-tetracyanopropene [32019-26-4], 1-propene-1,1,3,3-tetracarbonitrile (52); 1,1,2,3,3-pentacyanopropene [45078-17-9], 1-propene-1,1,2,3,3-pentacarbonitrile (51); 1,1,2,6,7,7-hexacyano-1,3,5-heptatriene [69239-39-0] (53); 2-dicyanomethylene-1,1,3,3-tetracyanopropane [32019-27-5] (51); and 1,3-cyclopentadiene-1,2,3,4,5-pentacarbonitrile [69239-40-3] (54,55). Many of these acids rival mineral acids in strength (56) and are usually isolable only as salts with metal or ammonium ions. The remarkable strength of these acids results from resonance stabilization in the anions that is not possible in the protonated forms.

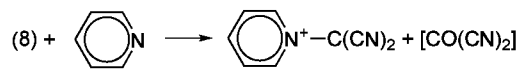
Some of the cyanocarbon anions are highly colored, eg, salts of 1,1,2,6,7,7-hexacyano-1,3,5-heptatriene are deep blue, $\log_{10} \epsilon = 4.22$ at 635 nm (53).

Oxacyanocarbons

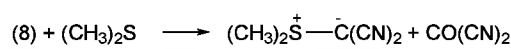
Tetracyanoethylene oxide [3189-43-3] (8), oxiranetetracarboxitrile, is the most notable member of the class of oxacyanocarbons (57). It is made by treating TCNE with hydrogen peroxide in acetonitrile. In reactions unprecedented for olefin oxides, it adds to olefins to form 2,2,5,5-tetracyanotetrahydrofuran [3041-31-4] in the case of ethylene, acetylenes, and aromatic hydrocarbons via cleavage of the ring C—C bond. The benzene adduct (9) is 3a,7a-dihydro-1,1,3,3-phthalantetracarboxitrile [3041-36-9], $C_{12}H_4N_4O$.



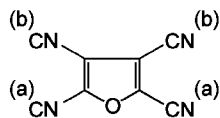
With pyridine, reaction takes place at the nitrogen rather than at a double bond, and an ylid [27032-01-5] is formed (57,58). Sulfides react similarly to give sulfilidenes and carbonyl cyanide (59).



This is the most convenient synthesis of carbonyl cyanide [1115-12-4]. The product of (8) and dimethyl sulfide is dimethyl sulfonium dicyanomethylide [5362-78-7].

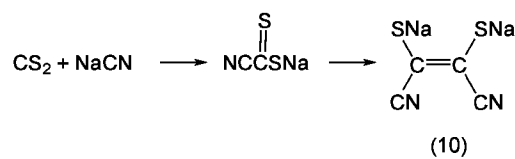


Tetracyanofuran [17989-87-6] 2,3,4,5-furantetracarbonitrile, has been obtained by dehydration of the corresponding tetracarboxamide (60). The α -cyano groups are more reactive, preferentially adding water, hydroxylamine, or phenylhydrazine.

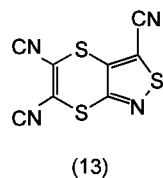
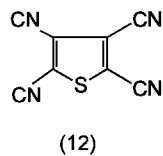
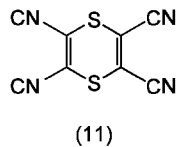


Thiacyanocarbons

For the most part, thiacyanocarbons are derived from "Bahr's Salt" [18820-77-4] (10), prepared from carbon disulfide and sodium cyanide (61) through the intermediacy of the sodium salt of carbonocyanidodithione acid [33498-03-2], C₂NS₂Na.

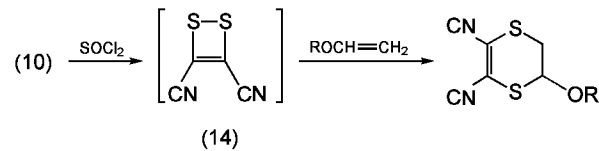


Oxidation of Bahr's Salt with iodine or thionyl chloride gives tetracyano-1,4-dithiin [2448-55-7] (11), 1,4-dithiin-2,3,5,6-tetracarbonitrile (62,63). The dithiin loses sulfur at 210°C to give tetracyanothiophene [4506-96-1] (12), 2,3,4,5-thio-phenetetracarbonitrile, and adds sulfur under the influence of sodium ethoxide to give the isothiazole [4656-27-3] (13) (63).



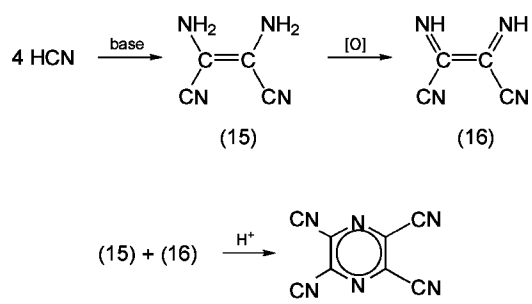
The nitrile groups of these three heterocyclic thiacyanocarbons behave normally and are convertible to amide, imide, or carboxylic acid groups.

Dicyano-1,2-dithiete [53562-16-6] (14) is thought to be an intermediate when Bahr's Salt is oxidized (64,65). If the oxidation is carried out in the presence of vinyl ethers, dihydrodithiins can be obtained in yields up to 60%.

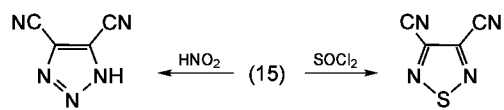


Azacyanocarbons

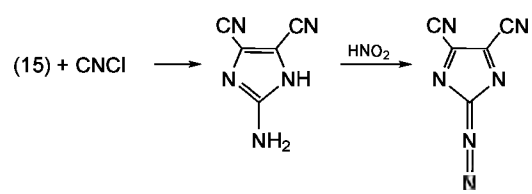
Hydrogen cyanide tetramer (*Z*-) 2,3-diamino-2-butenedinitrile [1187-42-4] (15), an azacyanocarbon, is produced by Nippon Soda in pilot-plant quantities for development as a chemical intermediate (66,67). On oxidation it forms 2,3-diiminobutanedinitrile [28321-79-7] (16) (68). These two, in turn, combine to give pyrazine–tetracarboxynitrile [33420-37-0] (69).



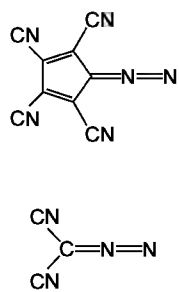
Treatment of (15) with nitrous acid gives 1*H*-1,2,3-triazole-4,5-dicarbonitrile [53817-16-6] (70). The action of thionyl chloride on (15) gives 1,2,5-thiadiazole-3,4-dicarbonitrile [23347-22-0] (71).



Reaction of (15) with cyanogen chloride gives 2-amino-1*H*-imidazole-4,5-dicarbonitrile [40953-34-2] (72). This amino compound reacts with nitrous acid to form 2-diazo-1*H*-imidazole-4,5-dicarbonitrile [51285-29-1] (72). Thermolysis of this diazo compound generates an extremely electrophilic carbene (72).



5-Diazo-1,3-cyclopentadiene-1,2,3,4-tetracarboxynitrile [54747-37-4] (54) and 2-diazopropanedinitrile [1618-08-2] (73) also possess reactive diazo groups.



Other azacyanocarbons are shown in Figure 1.

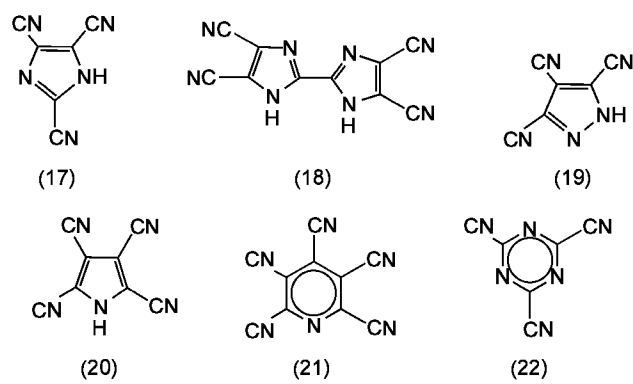


Fig. 1. Azacyanocarbons: 1*H*-imidazole-2,4,5-tricarbonitrile [69239-41-4] (17) (74,75); 4,4',5,5'-tetracyanobiimidazole [83312-48-7] (18) (76); 1*H*-pyrazole-3,4,5-tricarbonitrile [69239-42-5] (19) (77); 1*H*-pyrrole-2,3,4,5-tetracarboxynitrile [5231-17-4] (20) (78); 2,3,4,5,6-pyridinepentacarboxynitrile [69239-43-6] (21) (79); and 1,3,5-triazine-2,4,6-tricarbonitrile [7615-57-8] (22) (80).

Health and Safety Factors (Toxicology)

Unless specifically tested, all cyanocarbons should be considered as toxic as sodium cyanide or hydrogen cyanide (see CYANIDES). They should be used only in a fume hood, and rubber gloves should be worn.

BIBLIOGRAPHY

"Cyanocarbons" in *ECT* 2nd ed., Vol. 6, pp. 625–633, by B. C. McKusick and T. L. Cairns, E. I. du Pont de Nemours & Co., Inc.; in *ECT* 3rd ed., Vol. 7, pp. 359–369, by O. W. Webster, E. I. du Pont de Nemours & Co., Inc.

1. T. L. Cairns and co-workers, *J. Am. Chem. Soc.* **79**, 2340 (1957); *J. Am. Chem. Soc.* **80**, 2775 (1958).
2. **U.S. Pat. 3,166,584** (Jan. 19, 1965), T. L. Cairns and E. A. Graef (to E. I. du Pont de Nemours & Co., Inc.).
3. C. Moureu and J. Bongrand, *Compt. Rend. Acad. Sci.* **170**, 1025 (1920).
4. F. J. Brockman, *Can. J. Chem.* **33**, 507 (1955).
5. E. Cox and A. Fontaine, *Bull. Soc. Chim. Fr.*, 948 (1954); S. Trofimenko, *J. Org. Chem.* **28**, 2755 (1963).
6. C. E. Looney and J. R. Downing, *J. Am. Chem. Soc.* **80**, 2840 (1958).
7. D. A. Bekoe and K. N. Trueblood, *Z. Kristallogr.* **113**, 1 (1960).
8. R. E. Merrifield and W. D. Phillips, *J. Am. Chem. Soc.* **80**, 2778 (1958); G. D. Farnum, E. R. Atkinson, and W. C. Lothrop, *J. Org. Chem.* **26**, 3204 (1961); M. J. S. Dewar and H. Rogers, *J. Am. Chem. Soc.* **84**, 395 (1962).
9. R. A. Carboni, *Org. Synth.* **39**, 64 (1959).
10. **U.S. Pat. 3,118,929** (Jan. 21, 1964), E. L. Martin (to E. I. du Pont de Nemours & Co., Inc.).
11. O. W. Webster, W. Mahler, and R. E. Benson, *J. Am. Chem. Soc.* **84**, 3678 (1962).
12. J. M. Manriquez, G. T. Yee, R. S. McLean, A. J. Epstein, and J. S. Miller, *Science* **252**, 1415 (1991).
13. W. D. Phillips, J. C. Rowell, and S. I. Weissman, *J. Chem. Phys.* **33**, 626 (1960); P. H. Rieger and co-workers, *J. Am. Chem. Soc.* **85**, 683 (1963).
14. W. J. Middleton and co-workers, *J. Am. Chem. Soc.* **80**, 2783 (1958).
15. J. Sauer, *Angew. Chem.* **73**, 545 (1961).
16. M. Ozolins and G. H. Schenk, *Anal. Chem.* **33**, 1035 (1961); G. N. Schrauzer and S. Eichler, *Angew. Chem. Int. Ed. Engl.* **1**, 454 (1962); H. Prinzbach, *Angew. Chem.* **73**, 169 (1961); W. E. Doering and D. W. Wiley, *Tetrahedron* **11**, 183 (1960); K. Hafner and J. Schneider, *Justus Liebigs Ann. Chem.* **624**, 37 (1959); R. Criegie, *Angew. Chem. Int. Ed. Engl.* **1**, 519 (1962). These references contain examples of the use of TCNE to determine or characterize dienes.
17. A. T. Blomquist and Y. C. Meinwald, *J. Am. Chem. Soc.* **81**, 667 (1959); J. K. Williams, D. W. Wiley, and B. C. McKusick, *J. Am. Chem. Soc.* **84**, 2210 (1962); C. A. Stewart, *J. Org. Chem.* **28**, 3320 (1963).
18. W. J. Middleton and co-workers, *J. Am. Chem. Soc.* **80**, 2795 (1958).
19. C. L. Dickinson, D. W. Wiley, and B. C. McKusick, *J. Am. Chem. Soc.* **82**, 6132 (1960).
20. W. J. Middleton and V. A. Engelhardt, *J. Am. Chem. Soc.* **80**, 2788 (1958).
21. B. C. McKusick and co-workers, *J. Am. Chem. Soc.* **80**, 2806 (1958).
22. Z. Rappoport, *J. Chem. Soc.*, 4498 (1963); P. G. Farrell and J. Newton, *Tetrahedron Lett.*, 189 (1964).
23. G. N. Sausen, V. A. Engelhardt, and W. J. Middleton, *J. Am. Chem. Soc.* **80**, 2815 (1958).
24. B. Smith and U. Persmark, *Acta Chem. Scand.* **17**, 651 (1963).
25. W. E. Noland, W. C. Kuryla, and R. F. Lange, *J. Am. Chem. Soc.* **81**, 6010 (1959).
26. J. R. Roland and B. C. McKusick, *J. Am. Chem. Soc.* **83**, 1652 (1961).
27. K. Hafner and K. Moritz, *Justus Liebigs Ann. Chem.* **650**, 92 (1961).
28. D. J. Cram and R. D. Partos, *J. Am. Chem. Soc.* **85**, 1273 (1963).
29. Int. Pat. App. WO88 09350 (Dec. 1, 1988), R. P. Cope (to Crest-Foam Corp.).
30. O. W. Webster, *J. Am. Chem. Soc.* **84**, 3370 (1962).
31. D. S. Acker and W. R. Hertler, *J. Am. Chem. Soc.* **84**, 3370 (1962).
32. J. Diekmann and C. J. Pedersen, *J. Org. Chem.* **28**, 2879 (1963).
33. W. R. Hertler and co-workers, *J. Am. Chem. Soc.* **84**, 3387 (1962).
34. J. K. Williams, *J. Am. Chem. Soc.* **84**, 3478 (1962).
35. L. R. Melby and co-workers, *J. Am. Chem. Soc.* **84**, 3374 (1962).
36. P. H. H. Fischer and C. A. McDowell, *J. Am. Chem. Soc.* **85**, 2694 (1963).
37. W. J. Siemons, P. E. Bierstedt, and R. G. Kepler, *J. Chem. Phys.* **39**, 3523 (1963).
38. J. Ferraris and co-workers, *J. Am. Chem. Soc.* **95**, 948 (1973).
39. L. B. Coleman and co-workers, *Solid State Commun.* **12**, 1125 (1973).
40. M. J. Cohen and co-workers, *Phys. Rev. Sect. B.* **13**, 5111 (1976).
41. G. A. Thomas and co-workers, *Phys. Rev. Sect. B.* **13**, 5105 (1976).
42. E. Mayer, *Monatsh. Chem.* **100**, 462 (1969).
43. S. Trofimenko and B. C. McKusick, *J. Am. Chem. Soc.* **84**, 3677 (1962).
44. A. T. Blomquist and E. C. Winslow, *J. Org. Chem.* **10**, 149 (1945).
45. A. D. Kirshenbaum and A. V. Grosse, *J. Am. Chem. Soc.* **75**, 499 (1953).
46. A. W. Johnson, *The Chemistry of the Acetylenic Compounds*, Vol. II, Longmans, Green & Co., New York, 1950, pp. 258–260.
47. E. Ciganek, *Tetrahedron Lett.*, 3321 (1967).
48. K. Wallenfels and K. Friedrich, *Tetrahedron Lett.*, 1223 (1963).
49. K. Friedrich, *Chem. Ber.* **103**, 3951 (1970).
50. K. Wallenfels and G. Bachmann, *Angew. Chem.* **73**, 142 (1961); K. Wallenfels and co-workers, *Tetrahedron* **21**, 2239 (1965); U.S. Pat. 3,114,756 (Dec. 17, 1963) (to Farbwerke Hoechst).
51. W. J. Middleton and co-workers, *J. Am. Chem. Soc.* **80**, 2795 (1958).
52. Y. Urushibara, *Bull. Chem. Soc. Jpn.* **2**, 278 (1927).
53. J. K. Williams, D. W. Wiley, and B. C. McKusick, *J. Am. Chem. Soc.* **84**, 2216 (1962).
54. O. W. Webster, *J. Am. Chem. Soc.* **88**, 4055 (1966).
55. *Ibid.*, 3046 (1966).
56. R. H. Boyd, *J. Phys. Chem.* **67**, 737 (1963).
57. W. J. Linn, O. W. Webster, and R. E. Benson, *J. Am. Chem. Soc.* **85**, 2032 (1963).
58. A. Rieche and P. Dietrich, *Chem. Ber.* **96**, 3044 (1963).
59. **U.S. Pat. 3,115,517** (Dec. 24, 1963), W. J. Linn (to E. I. du Pont de Nemours & Co., Inc.).
60. C. Weis, *J. Org. Chem.* **27**, 3514 (1962).
61. G. Bahr and G. Schleitzer, *Chem. Ber.* **88**, 1771 (1955); **90**, 438 (1957); G. Bahr, *Angew. Chem.* **68**, 525 (1956).
62. G. Bahr, *Angew. Chem.* **70**, 606 (1958).
63. H. E. Simmons and co-workers, *J. Am. Chem. Soc.* **84**, 4746 (1962).
64. H. E. Simmons, D. C. Blomstrom, and R. D. Vest, *J. Am. Chem. Soc.* **84**, 4756 (1962).
65. *Ibid.*, 4773, 4782 (1962).
66. H. Bredereck, G. Schmutzer, and E. Oehler, *Justus Liebigs Ann. Chem.* **600**, 81 (1956).
67. **Jpn. Pat. Appl. 61 134,362** (June 21, 1986), S. Sogabe (to Nippo Kagaku Kogyo Co., Ltd.).
68. O. W. Webster and co-workers, *J. Org. Chem.* **37**, 4133 (1972).
69. R. W. Begland and co-workers, *J. Org. Chem.* **39**, 1235 (1974).

70. H. Bredereck and G. Schmitzter, *Justus Liebigs Ann. Chem.* **600**, 95 (1956).
71. **U.S. Pats. 2,990,408** and **2,990,409** (June 27, 1961), M. Carmack, D. Shew, and L. M. Weinstock.
72. W. A. Sheppard and O. W. Webster, *J. Am. Chem. Soc.* **95**, 2695 (1973).
73. E. Ciganek, *J. Am. Chem. Soc.* **88**, 1979 (1966); **89**, 1454 (1967).
74. **U.S. Pat. 3,793,339** (Feb. 19, 1974), O. W. Webster (to E. I. du Pont de Nemours & Co., Inc.).
75. P. G. Apen and P. G. Rasmussen, *Heterocycles* **29**, 1325 (1989).
76. P. G. Rasmussen, R. L. Haugh, J. E. Anderson, O. H. Bailey, and J. C. Bayon, *J. Am. Chem. Soc.* **104**, 6155 (1982).
77. C. D. Weis, *J. Org. Chem.* **27**, 3695 (1962).
78. **U.S. Pat. 3,221,024** (Nov. 30, 1965), H. E. Simmons (to E. I. du Pont de Nemours & Co., Inc.).
79. P. Neumann, Ph.D. dissertation, University of Freiburg, Germany, 1967.
80. E. Ott, *Chem. Ber.* **52B**, 656 (1919).

General References

E. Ciganek, W. J. Linn, and O. W. Webster, in Z. Rapport, ed., *The Chemistry of the Cyano Group*, Interscience Publishers, London, 1970, pp. 423–638. An extensive review of cyanocarbon chemistry.
T. L. Cairns and B. C. McKusick, *Angew. Chem.* **73**, 520 (1961).
F. Freeman, *Chem. Revs.* **9**, 591 (1969). Review of malonitrile chemistry.
V. Migrdichian, *The Chemistry of Organic Cyanogen Compounds*, Reinhold Publishing Corp., New York, 1947.
K. Wallenfels and co-workers, *Angew. Chem. Int. Ed. Engl.* **15**, 261 (1976).
W. R. Hertler and co-workers, *Mol. Cryst. Liq. Cryst.* **171**, 205 (1989). Cyanocarbon history.

Owen W. Webster
E. I. du Pont de Nemours & Co., Inc.

CYANOHYDRINS

A cyanohydrin is an organic compound that contains both a cyanide and a hydroxy group on an aliphatic section of the molecule. Cyanohydrins are usually α -hydroxy nitriles which are the products of base-catalyzed addition of hydrogen cyanide to the carbonyl group of aldehydes and ketones. The IUPAC name for cyanohydrins is based on the α -hydroxy nitrile name. Common names of cyanohydrins are derived from the aldehyde or ketone from which they are formed (Table 1).

Table 1. Cyanohydrin Nomenclature

Name	CAS Registry Number	Formula	Synonyms
hydroxyacetonitrile	[107-16-4]	HOCH ₂ CN	formaldehyde cyanohydrin, glycolonitrile
2-hydroxypropanenitrile	[78-97-7]	HOHCN CH₃	acetaldehyde cyanohydrin, lactonitrile
2-hydroxy-2-methylpropanenitrile	[75-86-5]	OH CH₃CCN CH₃	acetone cyanohydrin, α -hydroxyisobutyronitrile, 2-methylactonitrile
1-hydroxy-cyclohexanecarbonitrile	[931-97-5]	OH CN 	cyclohexanone cyanohydrin
α -hydroxy-benzeneacetonitrile	[532-28-5]	OH -CHCN	benzaldehyde cyanohydrin, mandelonitrile
3-hydroxypropanenitrile ^a	[109-78-4]	HOCH ₂ CH ₂ CN	ethylene cyanohydrin, hydracrylonitrile
3-hydroxy-2-methylpropanenitrile ^a	[2567-01-3]	CH₃ HOCH₂CHCN	propylene cyanohydrin, 2-methylhydraacrylonitrile

^a A β -hydroxy nitrile.

The outstanding chemical property of cyanohydrins is the ready conversion to α -hydroxy acids and derivatives, especially α -amino and α,β -unsaturated acids. Because cyanohydrins are primarily used as chemical intermediates, data on production and prices are not usually published. The industrial significance of cyanohydrins is waning as more direct and efficient routes to the desired products are developed. Acetone cyanohydrin is the world's most prominent industrial cyanohydrin because it offers the main route to methyl methacrylate manufacture.

Physical Properties

Cyanohydrins are usually colorless to straw yellow liquids with an objectionable odor akin to that of hydrogen cyanide. The lower molecular-weight cyanohydrins can be distilled under reduced pressure provided the cyanohydrin is kept at a slightly acidic pH. Table 2 lists physical properties of some common cyanohydrins.

Table 2. Physical Properties of Some Cyanohydrins

Name	Mol wt	Mp, °C	Bp, ^a °C	Specific gravity	n_D^{20}	Flash point, °C
formaldehyde cyanohydrin	57.06	< -72	119 at 3.2 kPa ^b	1.104	1.4117	
acetaldehyde cyanohydrin	71.03	40	182–184, dec	0.988	1.4050	77
acetone cyanohydrin	85.10	19	85 at 3.1 kPa ^b	0.927	1.3992	74
cyclohexanone cyanohydrin	121.17	29	109–113 at 1.2 kPa ^b	1.032	1.4576	60
benzaldehyde cyanohydrin	133.15	10	170, dec	1.117	1.5315	
ethylene cyanohydrin	71.08	46.2	228	1.059	1.4256	129
propylene cyanohydrin	85.11		207		1.4280	

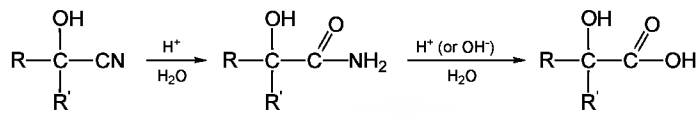
^a At 101.1 kPa ^b unless otherwise noted.

^b To convert kPa to mm Hg, multiply by 7.5.

Chemical Properties

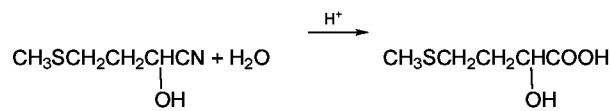
Cyanohydrins can react either at the nitrile group or at the hydroxyl group.

Nitrile Group. Hydrolysis of the nitrile group proceeds through the amide to the corresponding carboxylic acid. Because cyanohydrins are unstable at high pH, this hydrolysis must be catalyzed by acids. In cases where amide hydrolysis is slower than nitrile hydrolysis, the amide may be isolated.

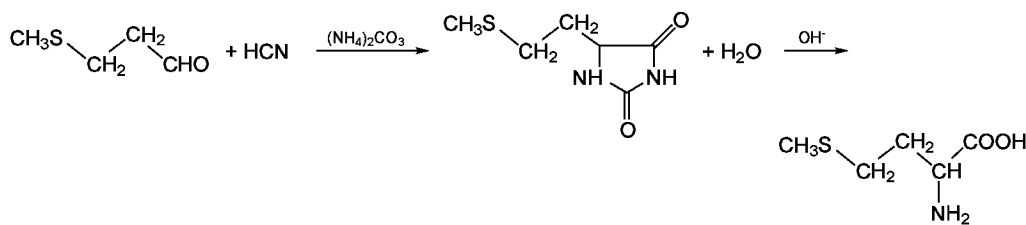


Thus acid hydrolysis of acetophenone cyanohydrin [20102-12-9] (R = C₆H₅, R' = CH₃) yields the corresponding amide which can be isolated. Further hydrolysis, usually with sodium hydroxide, gives atrolactic acid [575-30-0] in a 30% overall yield (1).

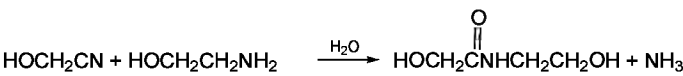
2-Hydroxy-4-methylthiobutyric acid [583-91-5], the hydroxy analogue of the amino acid methionine, is manufactured by acid hydrolysis of 3-methylthiopropionaldehyde cyanohydrin [17773-41-0], which is produced by the reaction of methyl mercaptan with acrolein (qv).



The mixture of D and L optical forms of this hydroxy analogue of methionine is converted to the calcium salt which is used in animal feed supplements. Cyanohydrins react with ammonium carbonate to form hydantoin (2), which yield amino acids upon hydrolysis. Commercial DL-methionine [59-57-8] is produced by hydrolysis of the hydantoin of 3-methylthiopropionaldehyde [3268-49-3].



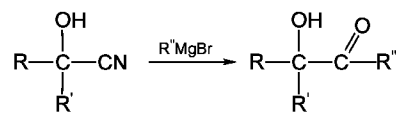
Reaction of cyanohydrins with absolute ethanol in the presence of HCl yields the ethyl esters of α-hydroxy acids (3). N-substituted amides can be synthesized by heating a cyanohydrin and an amine in water. Thus formaldehyde cyanohydrin and β-hydroxyethylamine lead to N-(β-hydroxyethyl)hydroxyacetamide (4).



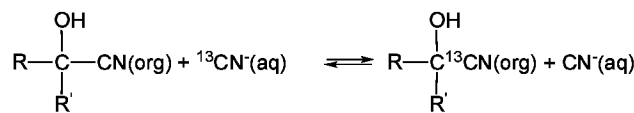
Catalytic hydrogenation of the nitrile function of cyanohydrins can give amines. As in the case of ordinary nitriles, catalytic reduction of cyanohydrins can yield a mixture of primary, secondary, and tertiary amines. Addition of acid or acetic anhydride to the reaction medium minimizes formation of secondary or tertiary amines through formation of the amine salt or acetamide derivative of the primary amine.

High yields of optically active cyanohydrins have been prepared from hydrogen cyanide and carbonyl compounds using an enzyme as catalyst. Reduction of these optically active cyanohydrins with lithium aluminum hydride in ether affords the corresponding substituted, optically active ethanolamine (5) (see ALKANOLAMINES).

Addition of hydrogen cyanide to an aldose to form a cyanohydrin is the first step in the Kiliani-Fischer method for increasing the carbon chain of aldoses by one unit. Cyanohydrins react with Grignard reagents (see GRIGNARD REACTION) to give α-hydroxy ketones.



Cyanide Exchange. Acetone cyanohydrin and methyl isobutyl ketone cyanohydrin [4131-68-4] dissolved in an organic solvent, such as diethyl ether or methyl isobutyl ketone, undergo cyanide exchange with aqueous cyanide ion to yield a significant cyanide carbon isotope separation. The two-phase system yields cyanohydrin enriched in carbon-13 and aqueous cyanide depleted in carbon-13. Equilibrium is obtained in seconds.



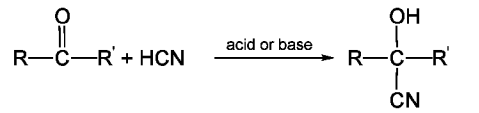
Some nitrogen isotope separation is also observed (6).

Hydroxyl Group. The OH group of cyanohydrins is subject to displacement with other electronegative groups. Cyanohydrins react with ammonia to yield amino nitriles. This is a step in the Strecker synthesis of amino acids. A one-step synthesis of α-amino acids involves treatment of cyanohydrins with ammonia and ammonium carbonate under pressure. Thus acetone cyanohydrin, when heated at 160°C with ammonia and ammonium carbonate for 6 h, gives α-aminoisobutyric acid [62-57-7] in 86% yield (7). Primary and secondary amines can also be used to displace the hydroxyl group to obtain N-substituted and N,N-disubstituted α-amino nitriles. The Strecker synthesis can also be applied to aromatic ketones. Similarly, hydrazine reacts with two molecules of cyanohydrin to give the disubstituted hydrazine.

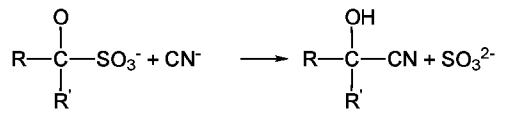
Cyanohydrins react with hydrogen halides or PCl₅ to give α-halo nitriles which can be further hydrolyzed to the α-halo carboxylic acids. The α-hydroxy group of cyanohydrins can be esterified with an acid or acid chloride. Dehydration of cyanohydrins with phosphorus pentoxide gives >80% yields of alkylacrylonitriles (8).

Preparation

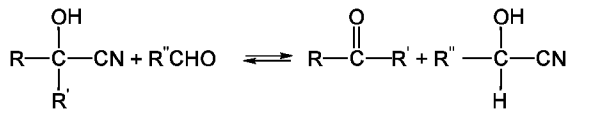
Cyanohydrins can be formed by (1) the acid- or base-catalyzed reaction of hydrogen cyanide with an aldehyde or ketone:



(2) the displacement of bisulfite ion by cyanide ion on the bisulfite addition compounds of aldehydes and ketones:

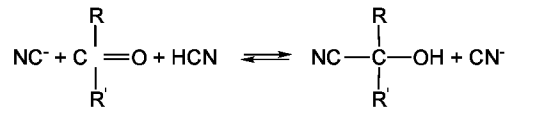


or (3) the exchange of cyanide ion between a ketone cyanohydrin and an aldehyde to give the usually more stable aldehyde cyanohydrin:



Direct combination of hydrogen cyanide and a carbonyl compound is the commercial and most common route to cyanohydrins.

The base-catalyzed addition of hydrogen cyanide to carbonyl compounds was one of the first mechanistic studies in organic chemistry (9). Addition of the cyanide ion to the carbonyl group is the rate controlling step. The reaction is slightly subject to general acid catalysis (10,11), an observation common to many other nucleophilic additions to carbonyl groups. Cyanohydrin formation probably proceeds by attack of cyanide ion on a carbonyl group that is already hydrogen bonded to a proton donor, ie, hydrogen cyanide (12).



The reaction is exothermic and reversible. Stabilization of the product mixture with acid is necessary before the cyanohydrin can be isolated, usually by distillation at reduced pressure.

All aliphatic aldehydes and most ketones react to form cyanohydrins. The lower reactivity of ketones, relative to aldehydes, is attributed to a combination of electron-donating effects and increased steric hindrance of the second alkyl group in the ketones. The magnitude of the equilibrium constants for the addition of hydrogen cyanide to a carbonyl group is a measure of the stability of the cyanohydrin relative to the carbonyl compound plus hydrogen cyanide:

$$K = \frac{[\text{cyanohydrin}]}{[\text{HCN}][\text{carbonyl compound}]}$$

The large values of the equilibrium constants for the aldehydes listed in Table 3 indicate the ease of forming aldehyde cyanohydrins relative to the majority of the ketone cyanohydrins. The most sterically hindered ketones, diisopropyl ketone and isopropyl *t*-butyl ketone, do not form appreciable amounts of cyanohydrins. Electron-deficient groups, which stabilize the electron-deficient carbonyl carbon atom, decrease *K*. Aromatic aldehydes form cyanohydrins, but an excess of aromatic aldehyde in basic cyanide solution can lead to further reaction to give benzoin condensation products. Alkyl aryl ketones can form cyanohydrins to a limited extent although the equilibrium usually falls far on the side of the reactants, ketone plus hydrogen cyanide. Diaryl ketones do not form cyanohydrins by this route. Alkyl aryl ketones can be converted to cyanohydrins by treatment with either diethyl aluminum cyanide (15), or by treating with (CH₃)₃SiCN followed by 3*N* HCl (16). The large cyanohydrin formation constants of some cyclohexanones listed in Table 3 are attributed to a reduction of steric strain of the six-membered ring of the cyanohydrin relative to that of the ketone (13). The stability of cyclohexanone cyanohydrin is such that treatment of the cyanohydrin with potassium hydroxide does not give HCN and ketone, but yields instead the potassium salt of the cyanohydrin.

Table 3. Equilibrium Constants for Formation of Cyanohydrins from Hydrogen Cyanide Plus Carbonyl Compounds^a

Carbonyl compound	<i>K</i> , M ⁻¹ b	References
acetaldehyde	7100	13
propionaldehyde	476	14
<i>n</i> -butyraldehyde	1042	14
isobutyraldehyde	1042	14
acetone	28	13
methyl ethyl ketone	33	13
methyl isopropyl ketone	52	13
methyl <i>t</i> -butyl ketone	29	13
isopropyl <i>t</i> -butyl ketone	0.12	13
methyl cyclohexyl ketone	56	13
cyclobutanone	108	13
cyclopentanone	34	13
cyclohexanone	1700	13
2-methylcyclohexanone	930	13
2- <i>t</i> -butylcyclohexanone	7.6	13
4- <i>t</i> -butylcyclohexanone	1700	13
cycloheptanone	7.7	14
benzaldehyde	200	13
<i>p</i> -chlorobenzaldehyde	204	14
<i>p</i> -nitrobenzaldehyde	55	14
<i>p</i> -methoxybenzaldehyde	32	14
<i>m</i> -chlorobenzaldehyde	400	14
<i>m</i> -nitrobenzaldehyde	370	14
<i>m</i> -methoxybenzaldehyde	233	14
<i>o</i> -chlorobenzaldehyde	1000	14
<i>o</i> -nitrobenzaldehyde	1429	14

<i>o</i> -methoxybenzaldehyde	385	14
acetophenone	0.8	14
ethyl phenyl ketone	1.7	14
<i>n</i> -propyl phenyl ketone	1.1	14
isopropyl phenyl ketone	4.0	14
isobutyl phenyl ketone	0.6	14
<i>t</i> -butyl phenyl ketone	11.1	14

^a Equilibrium constants measured at 20–25°C, except isopropyl *t*-butyl ketone which was measured at 35°C.

^b Reciprocal molarity or L. mol.

Production of cyanohydrins is accomplished through the base-catalyzed combination of hydrogen cyanide and the carbonyl compound in a solvent, usually the cyanohydrin itself (17). The reaction is carried out at high dilution of the feeds, at 10–15°C, and pH 6.5–7.5. The product is continuously removed from the reaction zone, cooled to push the equilibrium toward cyanohydrin formation, and then stabilized with mineral acid. Purification is usually effected by distillation.

Shipping, Storage, and Handling

Cyanohydrins should be stabilized with acid to pH 3–4 to prevent decomposition to hydrogen cyanide and carbonyl compound (18). Formaldehyde cyanohydrin should be considered an explosion and fire risk, particularly if impure. Acetone cyanohydrin and lactonitrile are usually manufactured and used at the same site without shipping. When shipped, steel drums, carboys, tank cars, and barges are used (19). In general, cyanohydrins are combustible liquids and many decompose upon heating. They should be stored in a cool, dry place preferably outside and separated from other storage. Containers should be protected against physical damage (19,20).

Health and Safety

Cyanohydrins are highly toxic by inhalation or ingestion, and moderately toxic through skin absorption (21). All α-hydroxy nitriles are potential sources of hydrogen cyanide or cyanides and must be handled with considerable caution. Contact with the skin and inhalation should be rigorously avoided. Special protective clothing should be worn and any exposure should be avoided (18,20). The area should be adequately ventilated. Immediate medical attention is essential in case of cyanohydrin poisoning.

Uses

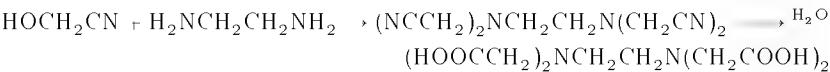
Cyanohydrins are used primarily as intermediates in the production of other chemicals. Manufacture of methyl methacrylate, used to make acrylic molding resins and clear sheet, eg, Plexiglas acrylic sheet, from acetone cyanohydrin is the most economically important cyanohydrin process (see METHACRYLIC POLYMERS). Cyanohydrins are also used as solvents in applications including fiber-spinning and metals refining. Cyanohydrins and derivatives reportedly act as antiknock agents in fuel oil and motor fuels and serve as electrolytes in electrolytic capacitors.

Specific Compounds

Formaldehyde Cyanohydrin. This cyanohydrin, also known as glycolonitrile [107-16-4], is a colorless liquid with a cyanide odor. It is soluble in water, alcohol, and diethyl ether. Equimolar amounts of 37% formaldehyde and aqueous hydrogen cyanide mixed with a sodium hydroxide catalyst at 2°C for one hour give formaldehyde cyanohydrin in 79.5% yield (22).

Although usually handled as an aqueous solution, formaldehyde cyanohydrin can be isolated in the anhydrous form by ether extraction, followed by drying and vacuum distillation (23). Pure formaldehyde cyanohydrin tends to be unstable especially at high pH. Small amounts of phosphoric acid or monochloroacetic acid are usually added as a stabilizer. Monochloroacetic acid is especially suited to this purpose because it codistills with formaldehyde cyanohydrin (24). Properly purified formaldehyde cyanohydrin has excellent stability (25).

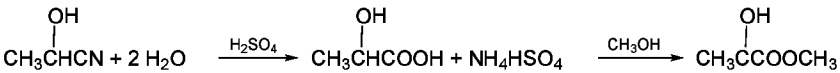
Direct reaction of formaldehyde cyanohydrin and ethylenediamine in the presence of a sulfuric acid catalyst gives ethylenediaminetetraacetonitrile [5766-67-6], hydrolysis of which leads to ethylenediaminetetraacetic acid [60-00-4] (EDTA), a widely used sequestering agent (26).



Nitrilotriacetonitrile [628-87-5], N(CH₂CN)₃, a precursor to nitrilotriacetic acid [139-13-9], N(CH₂COOH)₃, can be prepared from the reaction of formaldehyde cyanohydrin with ammonia (26). Formaldehyde cyanohydrin is also used as an intermediate in pharmaceutical production. Commercial formaldehyde cyanohydrin is available as a 70% aqueous solution stabilized by phosphoric acid.

Formaldehyde cyanohydrin was detected in Halley's comet by the Vega I space probe (27).

Acetaldehyde Cyanohydrin. This cyanohydrin, commonly known as lactonitrile, is soluble in water and alcohol, but insoluble in diethyl ether and carbon disulfide. Lactonitrile is used chiefly to manufacture lactic acid and its derivatives, primarily ethyl lactate. Lactonitrile [78-97-7] is manufactured from equimolar amounts of acetaldehyde and hydrogen cyanide containing 1.5% of 20% NaOH at 10–20 °C. The product is stabilized with sulfuric acid (28). Sulfuric acid hydrolyzes the nitrile to give a mixture of lactic acid [598-82-3] and ammonium bisulfate.

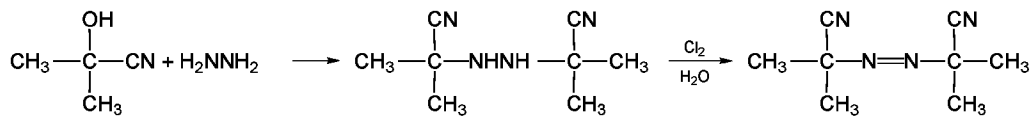


This mixture can be purified by adding methanol to form methyl lactate [547-64-8] which is separated from the ammonium bisulfate. The methyl lactate is distilled, then hydrolyzed back to the aqueous acid. Removal of most of the water yields 90% lactic acid (29).

Acetone Cyanohydrin. This cyanohydrin, also known as α-hydroxyisobutyronitrile and 2-methylactonitrile [75-86-5], is very soluble in water, diethyl ether, and alcohol, but only slightly soluble in carbon disulfide or petroleum ether. Acetone cyanohydrin is the most important commercial cyanohydrin as it offers the principal commercial route to methacrylic acid and its derivatives, mainly methyl methacrylate [80-62-6] (see METHACRYLIC ACID AND DERIVATIVES). The principal U.S. manufacturers are Rohm and Haas Co., Du Pont, CyRo Industries, and BP Chemicals. Production of acetone cyanohydrin in 1989 was 582,000 metric tons (30).

Acetone cyanohydrin is used as a raw material for insecticides manufacture and also to produce ethyl α-hydroxyisobutyrate [80-55-7], a pharmaceutical intermediate. It has been used as a complexing agent for metals refining and separation. Acetone cyanohydrin complexes can be used to separate Ni²⁺, Cu²⁺, Hg²⁺, Zn²⁺, Cd²⁺, or Fe²⁺ from Mg²⁺, Ba²⁺, Ca²⁺, Na⁺, or K⁺ on strongly basic ion-exchange resins (31). Acetone cyanohydrin is also used as a reagent in the formation of aldehyde cyanohydrins from aldehydes; in combination with a KCN-crown ether complex, it acts as an effective, stereoselective hydrocyanating reagent (32).

Reaction of hydrazine with acetone cyanohydrin gives a disubstituted hydrazine, which upon oxidation with chlorine water gives 2,2'-azobisisobutyronitrile [78-67-1] (AIBN), a stable, colorless, crystalline material at room temperature.



When heated to 120°C, AIBN decomposes to form nitrogen and two 2-cyanoisopropyl radicals. The ease with which AIBN forms radicals, and the fact that the rate of information does not vary much in various solvents has resulted in wide use of AIBN as a free-radical initiator. AIBN is used commercially as a catalyst for vinyl polymerization (see INITIATORS).

Acetone cyanohydrin is manufactured (33) by the direct reaction of hydrogen cyanide with acetone, catalyzed by base, generally in a continuous process (Fig. 1). Acetone and hydrogen cyanide are fed continuously with a small amount of sodium hydroxide catalyst to the stirred acetone cyanohydrin generator. Since the reaction is exothermic, acetone cyanohydrin is also used as the reaction solvent to aid in dissipating the heat of reaction. To push the equilibrium toward cyanohydrin formation, the reaction mixture is chilled first in the generator, then in the hold tank. The crude acetone cyanohydrin is neutralized with sulfuric acid to pH 1–2 (19). The acidification causes precipitation of the sodium catalyst as sodium sulfate salts which are then filtered. The crude acetone cyanohydrin is fed to a light-ends stripping column where the hydrogen cyanide, acetone, and some of the water are removed overhead and recycled to the acetone cyanohydrin generator. The concentrated acetone cyanohydrin bottoms are then sent to a dehydration column where the remainder of the water is removed under vacuum. The anhydrous acetone cyanohydrin is removed as the bottoms product and used directly in methacrylate manufacture.

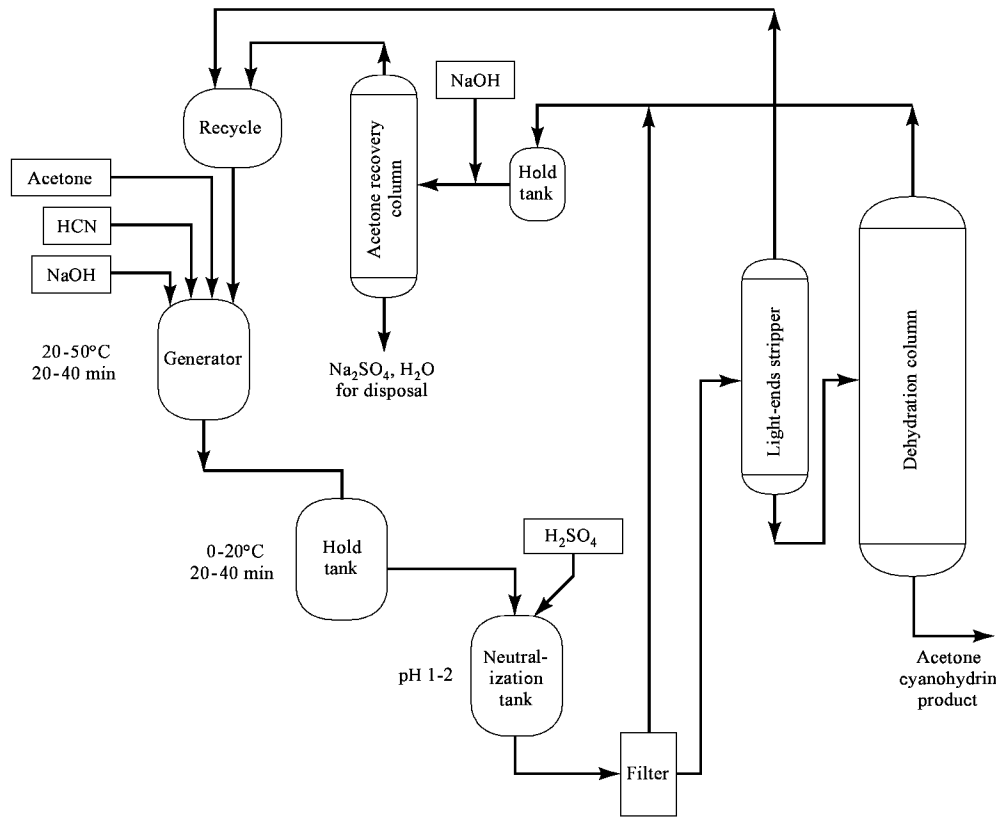
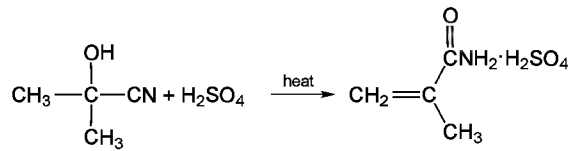


Fig. 1. Acetone cyanohydrin process.

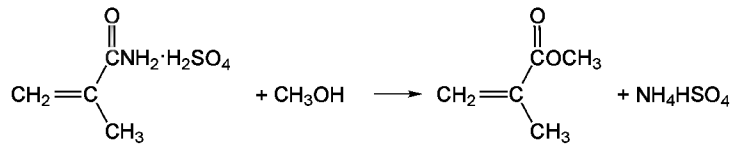
Single-pass conversions of acetone cyanohydrin are 90–95% depending on the residence times and temperatures in the generator and hold tank. Overall yields of product from acetone and hydrogen cyanide can be >97%. There are no significant by-products of the reaction other than the sodium salts produced by neutralization of the catalyst.

The hydrogen cyanide used in the process is sometimes produced on-site through catalytic conversion of an ammonia-methane-air mixture directly to hydrogen cyanide, the Andrussow process. However, by-product hydrogen cyanide from acrylonitrile (qv) manufacture is increasingly being used in acetone cyanohydrin production. Although sodium hydroxide was used as the catalyst in the above example, any of a number of other basic catalysts would suffice including KOH, K₂CO₃ (18), ion-exchange resins (34), and acetic acid-sodium acetate buffers (35).

Acetone cyanohydrin is used for methacrylate manufacture. Sulfuric acid is added in greater than an equimolar amount, and by thermal cracking the acetone cyanohydrin is converted to methacrylamide sulfate [29194-31-8].



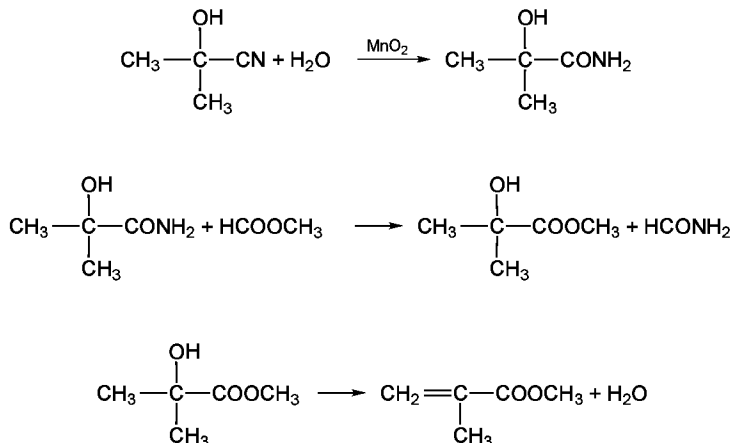
The methacrylamide sulfate is esterified with methanol to give methyl methacrylate and ammonium bisulfate [7802-63-6] as a by-product.



The basic process for manufacture of acetone cyanohydrin was developed in the 1930s by Imperial Chemical Industries and has been improved over the years by the producing companies.

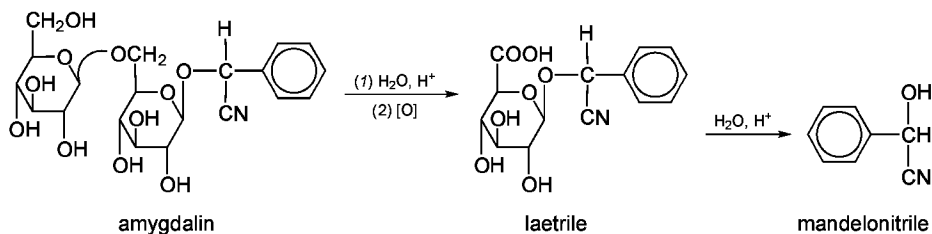
Conversion of acetone cyanohydrin to methyl methacrylate produces a large amount of ammonium bisulfate by-product which lacks ready marketability and is usually converted to sulfuric acid for reuse in the conversion of acetone cyanohydrin to methacrylates. The nitrogen values of the

ammonium bisulfate are lost in the recycle process. The necessity for the ammonium bisulfate recycle plants has created increased interest in routes to methacrylates that do not coproduce ammonium bisulfate (36). Mitsubishi Gas Chemical has developed a new route from acetone cyanohydrin to methyl methacrylate (37) which produces water as a by-product. In this process acetone cyanohydrin (α -hydroxyisobutyronitrile [75-86-5]) is hydrated to yield α -hydroxy isobutyramide [13027-88-8]. Ester interchange between α -hydroxy isobutyramide and methyl formate yields methyl α -hydroxy isobutyrate [2110-78-3] which is dehydrated to methyl methacrylate (MMA).



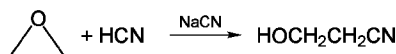
The formamide is dehydrated to HCN which is recycled back to make acetone cyanohydrin. The overall reaction is
acetone + methyl formate $\rightarrow$ MMA + H₂O.

Benzaldehyde Cyanohydrin. This cyanohydrin, also known as mandelonitrile [532-28-5], is a yellow, oily liquid, insoluble in water, but soluble in alcohol and diethyl ether. Mandelonitrile is a component of the glycoside amygdalin [29883-15-6], a precursor of laetrile [1332-94-7] found in the leaves and seeds on most *Prunus* species (plum, peach, apricot, etc). In 1832, mandelonitrile was the first cyanohydrin to be synthesized.



It is commercially prepared from benzaldehyde and hydrogen cyanide. Mandelonitrile is used by certain insects (tiger beetles, an African millipede) as a defense fluid (38). After expelling the fluid an enzyme catalyzes the conversion of mandelonitrile to benzaldehyde and HCN, which is usually fatal to the insect's enemy.

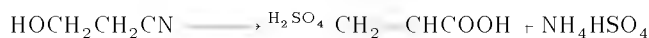
Ethylene Cyanohydrin. This cyanohydrin, also known as hydracrylonitrile or glycocyanohydrin [109-78-4], is a straw-colored liquid miscible with water, acetone, methyl ethyl ketone, and ethanol, and is insoluble in benzene, carbon disulfide, and carbon tetrachloride. Ethylene cyanohydrin differs from the other cyanohydrins discussed here in that it is a β -cyanohydrin. It is formed by the reaction of ethylene oxide with hydrogen cyanide.



Like the formation of α -cyanohydrins, this reaction is catalyzed by bases or cyanide ion, but unlike the α -cyanohydrin case this reaction is not reversible, and under certain conditions it can proceed with violence. Ethylene cyanohydrin can also be prepared by the reaction of ethylene chlorohydrin and alkali cyanides (39).

The first U.S. plant for acrylonitrile manufacture used an ethylene cyanohydrin feedstock. This was the primary route for acrylonitrile manufacture until the acetylene-based process began to replace it in 1953 (40). Maximum use of ethylene cyanohydrin to produce acrylonitrile occurred in 1963. Acrylonitrile (qv) has not been produced by this route since 1970.

The first commercial process for manufacture of acrylic acid (qv) and acrylates involved hydrolysis of ethylene cyanohydrin in aqueous sulfuric acid.



This route is no longer commercially significant.

BIBLIOGRAPHY

"Cyanohydrins" in *ECT* 1st ed., Vol. 4, pp. 755–759, by R. H. F. Manske, Dominican Rubber Co. Ltd. and H. S. Davis, American Cyanamid Co.; in *ECT* 2nd ed., Vol. 6, pp. 668–675, by A. P. Lurie, Eastman Kodak Co.; in *ECT* 3rd ed., Vol. 7, pp. 385–396, by M. S. Cholod, Rohm and Haas Co.

1. E. L. Eliel and J. P. Freeman, in N. Rabjohn, ed., *Organic Synthesis*, Coll. Vol. IV, John Wiley & Sons, Inc., New York, 1963, p. 58.
2. E. C. Wagner and M. Baizer, in E. C. Horning, ed., *Organic Syntheses*, Coll. Vol. III, John Wiley & Sons, Inc., New York, 1955, p. 323.
3. V. P. Velikov and co-workers, *Izv. Akad. Nauk SSSR Ser. Khim.* **8**, 1862 (1967).
4. **U.S. Pat. 3,190,916** (June 22, 1965), N. B. Rainer (to Coastal Interchemical Co.).
5. W. Becker, H. Freund, and E. Pfeil, *Angew. Chem. Int. Ed. Engl.* **4**(12), 1079 (1965).
6. L. L. Brown and J. S. Drury, *J. Inorg. Nucl. Chem.* **35**, 2897 (1973); U.S. Pat. 3,607,010 (Sept. 21, 1971), L. L. Brown (to U.S. Atomic Energy Commission).
7. **Neth. Appl. 6,607,754** (Dec. 5, 1966) (to Rnhm, GmbH).
8. S. I. Mekhtiev and R. G. Mamedov, *Azerb. Khim. Zh.* **2**, 110 (1971).
9. A. Lapworth, *J. Chem. Soc.* **83**, 995 (1903); A. Lapworth and R. H. F. Manske, *J. Chem. Soc.*, 2533 (1928); 1976 (1930).
10. W. J. Svirebely and J. F. Roth, *J. Am. Chem. Soc.* **75**, 3106 (1953).
11. J. Hine, *Physical Organic Chemistry*, McGraw-Hill Book Co., New York, 1956, p. 251.
12. H. H. Hustedt and E. Pfeil, *Justus Liebigs Ann. Chem.* **640**, 15 (1961).
13. J. Hine, *Structural Effects on Equilibria in Organic Chemistry*, John Wiley & Sons, Inc., New York, 1975, p. 259, and references cited therein.
14. V. Migrdichian, *The Chemistry of Organic Cyanogen Compounds*, Reinhold Publishing Co., New York, 1947, Chapt. 9.

15. W. Nagata, M. Yoshioka, and M. Murakami, *Org. Synth.* **52**, 96–99 (1972).

16. P. G. Gassman and J. J. Talley, *Tetrahedron Lett.* **40**, 3773–3778 (1978).

17. **Ger. Pat. 1,087,122** (Aug. 18, 1960), K. Sennewald and co-workers (to Knapsack-Grieshein Akt. Ges.).

18. **U.S. Pat. 2,537,814** (Jan. 9, 1957), H. S. Davis (to American Cyanamid Co.).

19. *Fire Protection Guide on Hazardous Materials*, 6th ed., National Fire Protection Association, Quincy, Mass., 1975.

20. N. I. Sax, ed., *Dangerous Properties of Industrial Materials*, 4th ed., Van Nostrand Reinhold Co., New York, 1975; *Toxic and Hazardous Industrial Chemicals Safety Manual*, The International Technical Information Institute, Tokyo, Japan, 1976.

21. H. E. Christensen, ed., *Registry of Toxic Effects of Chemical Substances*, Coll. Vol. III, John Wiley & Sons, Inc., New York, 1955, p. 436.

22. **U.S. Pat. 2,890,238** (June 9, 1959), A. R. Sexton (to The Dow Chemical Company).

23. R. Gaudry, in Ref. 2, p. 436.

24. **U.S. Pat. 2,623,896** (Dec. 30, 1952), H. Beier (to Rohm, GmbH).

25. **U.S. Pat. 3,057,903** (Oct. 9, 1962), J. W. Nemecek and C. H. McKeever (to Rohm and Haas Co.).

26. **U.S. Pat. 2,855,428** (Oct. 7, 1958), J. J. Singer and M. Weisberg (to Hampshire Chemical Co.).

27. J. Kissel and F. R. Krueger, *Nature (London)* **326**(6115), 755–760 (1987).

28. **Jpn. Pat. 68 29,576** (Dec. 18, 1968), S. Yamaguchi and co-workers (to Mitsubishi Chem. Ind. Co. Ltd.).

29. *Hydrocarbon Process.*, 156 (Nov. 1975); *Jpn. Pat.* 65 2333 (Feb. 6, 1965), T. Komada and co-workers (to Toyama Chem. Ind. Co. Ltd.).

30. *Synthetic Organic Chemicals, U.S. Production and Sales, 1989*, USITC Publication 2338, USITC, Dec. 1990, section 15-4.

31. L. Legradi, *J. Chromatogr.* **102**, 319 (1974).

32. C. L. Liotta, A. M. Dabdoub, and L. H. Zalkow, *Tetrahedron Lett.* **13**, 1117 (1977).

33. **U.S. Pat. 4,130,580** (Dec. 19, 1978), M. S. Cholod (to Rohm and Haas Co.).

34. M. Borreal and J. Modiano, *Chim. Ind. Paris* **78**, 632 (1957).

35. **U.S. Pat. 2,748,154** (May 29, 1956), G. E. Journeay (to Monsanto Chemical Co.).

36. Y. Oda and co-workers, *Hydrocarbon Process.*, 115 (Oct. 1975), for an example of an alternative route.

37. **U.S. Pat. 4,613,684** (May 9, 1985), T. Aoyama, K. Kida, and T. Uchiyama (to Mitsubishi Gas Chemical Co.); **EP 342,458** (May 5, 1989), I. Hideo and co-workers (to Mitsubishi Gas Chemical Co.); **U.S. Pat. 4,693,887** (Jan. 29, 1987), U. Von Oehsen and co-workers (to BASF A.-G.); *Chem. Week* **145**(4), 38 (July 26, 1989).

38. M.S. Blum and co-workers, *Comp. Biochem. Physiol. B* **69B**(4), 903–904 (1981).

39. E. C. Kendall and B. McKenzie, in H. Gilman, ed., *Organic Synthesis*, Coll. Vol. I, John Wiley & Sons, Inc., New York, 1932, p. 256.

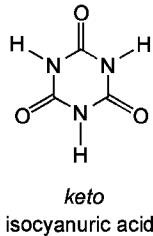
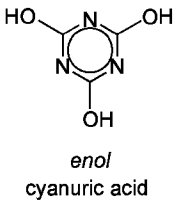
40. M. Sittig, *Acrylonitrile*, Noyes Development Corp., Park Ridge, N.J., 1965, p. 56.

Michael S. Cholod
Rohm and Haas Company

CYANURIC AND ISOCYANURIC ACIDS

Cyanuric acid [108-80-5] was discovered by Scheele in 1776 but achieved commercial importance only in the mid 1950s primarily via its derivatives. The chemistry of cyanuric is diversified because of multiple reaction sites. *N*-Chlorination of cyanuric acid produces chloroisocyanurates that are widely used as disinfectants, sanitizers, and bleaches. The triallyl- and tris(2-hydroxyethyl) derivatives are employed as cross-linking and curing agents, and tris(2,3-epoxypropyl) isocyanurate is used in weather-resistant powder coatings (qv). Melamine cyanurate finds significant use as a fire retardant in plastics. Considerable interest has developed in the use of cyanuric acid for reduction of nitrogen oxides in exhaust gases from combustion of oil, gas, and coal.

Nomenclature is based on the keto-enol tautomers. The trihydroxy form is variously designated cyanuric acid, *s*-triazine-2,4,6-triol or 2,4,6-trihydroxy-*s*-triazine. The trioxo structure, or *s*-triazine-2,4,6(1*H*,3*H*,5*H*)-trione is the basis for the isocyanuric acid nomenclature.



The triazine-based system is preferred in the scientific literature and is used by government regulatory agencies; however, the less cumbersome (iso)cyanurate designations persist in commerce. The keto form normally predominates in cases of keto–enol tautomerism. Indeed, current evidence based on ir, Raman, and uv spectroscopic, and x-ray crystallographic data suggests that the trioxo species prevails in the solid state and in solution (1–3). In alkaline solution, the anion formed is that of the hydroxy tautomer. Through common usage, both forms are collectively called cyanuric acid (CA).

Properties

Cyanuric acid is an odorless, white, crystalline solid that does not melt up to 330°C; at higher temperatures it sublimes and dissociates to isocyanic acid (HNCO) [75-13-8].



The vapor pressure of CA at 167–200°C is given by: $\log(\text{kPa}) = 5552/T + 11.54$ (*T* in K) (4). Over the 295–360°C range the vapor pressure is given by: $\log P(\text{kPa}) = 6740/T + 13.38$ (5). The equilibrium constants for dissociation of CA in the gas phase are 0.02 and 0.76 MPa² (1.9 and 74.3 atm²) at 365 and 434°C, respectively. Below 350°C, isocyanic acid tends to polymerize, whereas above 350°C it can form by depolymerization of CA. Isocyanic acid formed by thermal depolymerization of CA can be condensed at low temperature and employed in synthesis (6,7). It polymerizes at room temperature to a mixture of a trimer (HNCO)₃, CA and linear polymer (HNCO)_∞, cyamelide [462-02-2] (8). Cyamelide is converted to CA by heating at 200°C (9).

Cyanuric acid is a titrable weak acid (p*K*_{a1} = 6.88, p*K*_{a2} = 11.40, p*K*_{a3} = 13.5) (10). The pH of a saturated aqueous solution of pure CA at room temperature is ~4.8. Thermodynamic properties of CA are given in Table 1. Spectroscopic data are available (1–3). Proton nmr is of limited usefulness because of proton exchange and CA's symmetry and low solubility. Nuclear quadrupole resonance measurements (¹⁴N) have been reported (12).

Table 1. Thermodynamic Properties of CA^{a, b}

Property	Value	References
ΔH_f° , kJ/mol	690.8	
	660.2	18
ΔH_{comb} , kJ/mol	918.4	
ΔH_{subl} , kJ/mol	160.2	
at 25°C	133.6	4
at 295–360°C	128.9	5
<i>C_p</i> , J/(molK)	132.8	
ΔH_{neut} (NaOH), kJ/mol		
1st H	28.2	
2nd H	17.2	

^a Ref. 11 unless otherwise noted.

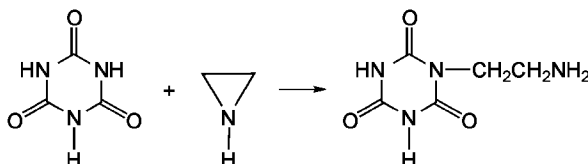
^b To convert J to cal, divide by 4.184.

Cyanuric acid is only slightly soluble (<0.1%) at room temperature in common organic solvents such as acetone, benzene, diethyl ether, ethanol, and hexane (13). Solubility is significant in basic nitrogen compounds (eg, dimethylformamide 7.2° o) or unusual solvents such as DMSO (17.4° o). Solubility in

dichloroisocyanurate double salts have been prepared (41). Some properties of CA and chloroisocyanurates are given in Reference 42.

Bromo and bromochloro derivatives of CA include $\text{HBrCl}(\text{NCO})_3$ [889325-49-2], $\text{HBr}_2(\text{NCO})_3$ [15114-43-9], $\text{NaBrCl}(\text{NCO})_3$ [20367-88-8], $\text{NaBr}_2(\text{NCO})_3$ [15114-43-8], $\text{KBr}_2(\text{NCO})_3$ [15114-46-2], $\text{BrCl}_2(\text{NCO})_3$ [89696-38-8], $\text{Br}_2\text{Cl}(\text{NCO})_3$, and $\text{Br}_3(\text{NCO})_3$ [17497-85-7] (43–45).

Organic Derivatives. Although numerous mono-, di-, and trisubstituted organic derivatives of cyanuric and isocyanuric acids appear in the literature, many are not accessible via cyanuric acid. Cyanuric chloride 2,4,6-trichloro-*s*-triazine [108-77-0], is generally employed as the intermediate to most cyanurates. Trisubstituted isocyanurates can also be produced by trimerization of either aliphatic or aromatic isocyanates with appropriate catalysts (46) (see ISOCYANATES, ORGANIC). Alkylation of CA generally produces trisubstituted isocyanurates even when a deliberate attempt is made to produce mono- or disubstituted derivatives. There are exceptions, as in the production of mono-2-aminoethyl isocyanurate [18503-66-7] in nearly quantitative yield by reaction of CA and aziridine in DMF (47).



Virtually all of the organo derivatives of CA are produced by reactions characteristic of a cyclic imide, wherein isocyanurate nitrogen (frequently as the anion) nucleophilically attacks a positively polarized carbon of the second reactant. Cyanuric acid and ethylene oxide react nearly quantitatively at 100°C to form tris(2-hydroxyethyl)isocyanurate [839-90-7] (THEIC) (48–52). Substitution of propylene oxide yields the hydroxypropyl analogue (48,49). At elevated temperatures (~200°C). CA and alkylene oxides react in inert solvent to give *N*-hydroxyalkyloxazolidones in approximately 70% yield (53). Alternatively, THEIC can be prepared by reaction of CA and 2-chloroethanol in aqueous caustic (52). THEIC can react further via its hydroxyl functionality to form esters, ethers, urethanes, phosphites, etc (54). Reaction of CA with epichlorohydrin in alkaline dioxane solution gives tris(3-chloro-2-hydroxypropyl)isocyanurate [7423-53-2] (55) (see CHLOROHYDRINS). This material can be converted into the commercial product, tris(2,3-epoxypropyl)isocyanurate [2451-62-9] (triglycidyl isocyanurate) by dehydrohalogenation with alkali (55). Reaction of CA with glycidol [556-52-5] yields tris(2,3-dihydroxypropyl)isocyanurate (56).

In a reaction analogous to the nucleophilic attack of cyanurate on 2-chloroethanol, the commercial product triallyl isocyanurate [1025-15-6] (TAIC) is formed in 82% yield by action of CA on allyl chloride at 130°C in *o*-dichlorobenzene with triethylamine as proton acceptor (49,57). Similarly, by substituting the appropriate organo halide, the following trisubstituted isocyanurates can be synthesized: tribenzyl isocyanurate (92% yield) [606-03-1] (57); trimethylallyl isocyanurate [6291-95-8] (89% yield) (49,57); tris(2-ketopropyl)isocyanurate [61050-97-3] (78% yield) (13); and tris(carbomethoxymethyl)isocyanurate [69455-18-1] (94% yield) (49,13). The last compound upon acidic hydrolysis has been reported to yield tris(carboxymethyl)isocyanurate [1968-52-1]. By contrast, ring rupture occurs under alkaline conditions (13,49,58). A mixed isocyanurate having both hydroxyethyl and allyl substituents is obtainable by the simultaneous action of 2-chloroethanol and allyl chloride on an alkaline solution of CA (59). Tris(isocyanatoethyl)isocyanurate is produced by reaction of CA with 1-chlorohexamethylene-6-isocyanate in the presence of triethylamine (60). Reaction of CA with propargyl bromide and caustic yields tris(propargyl)isocyanurate (61). Other compounds prepared by reactions of CA and organohalogen compounds include tris(carbamoylmethyl)isocyanurate [1843-48-7] from 2-chloroacetamide in 80% yield (58) and tri-*n*-hexyl isocyanurate [36761-61-2] from *n*-hexyl chloride in 71% yield (62).

Trimethyl isocyanurate [827-16-7] can be synthesized in 60% yield by the reaction of CA with dimethyl sulfate in alkaline medium (13) or with diazomethane (63); and in essentially quantitative yield by thermal rearrangement of trimethyl cyanurate [877-89-4]. Isomerization of alkyl cyanurates to the corresponding isocyanurates is frequently observed (11,64).

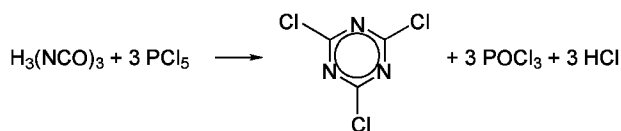
Cyanuric acid readily dissolves in aqueous formaldehyde forming tris(hydroxymethyl)isocyanurate [10471-40-6] (THMIC) which can be isolated by evaporation (11). THMIC in turn reacts with acetic anhydride to yield tris(acetoxymethyl)isocyanurate [54635-07-3], either thionyl chloride or phosphorus pentachloride to give tris(chloromethyl)isocyanurate [63579-00-0], and phenyl isocyanate in pyridine to yield tris(*N*-phenylcarbamoxymethyl) isocyanurate [21253-39-4] in 87% yield (65). Reaction of CA with paraformaldehyde and 2,6-dicyclohexylphenol yields tris(3,5-dicyclohexyl-4-hydroxybenzyl)isocyanurate (66). Cyanuric acid reacts with bicyclic amide acetals forming triols, useful for preparation of polyurethanes (67). Ketene in acetone reacts with CA under alkaline catalysis to give triacetyl cyanurate [13483-16-4] in high yield rather than the isocyanurate (68). *N,N',N''*-Triacetyl isocyanurate [35843-57-3] has been prepared by reaction of *N,N',N''*-tris(tributylstannyl)isocyanurate [752-58-9] and acetyl chloride (69).

Cyanuric acid adds across the activated double bond of acrylonitrile in alkaline dimethylformamide solution at 130°C. Both tris(2-cyanoethyl) [2904-28-1] (49) and bis(2-cyanoethyl)isocyanurate [2904-27-0] can be isolated; the trisubstituted derivative is obtained by precipitation, whereas the disubstituted compound remains in solution. Each readily undergoes acidic hydrolysis to form the corresponding tris and bis(2-carboxyethyl)isocyanurate [2904-40-7]. Tris(2-carboxyethyl)isocyanurate [2904-41-8] is easily converted to the ethyl ester, tris(2-carbomethoxyethyl)isocyanurate [2904-39-4] (49). Interestingly, hydrogenation of the tris(cyanoethyl) derivative in the presence of ammonia produces bis- [69455-19-2] and mono-3-aminopropyl [69455-20-5] isocyanurates (49). Cyanuric acid adds across the double bond of *N*-substituted maleimides forming trisubstituted isocyanurates which can be hydrolyzed, forming carboxylic acid and secondary amine functionalities (50). Addition of CA to acrolein produces tris(3-oxopropyl)isocyanurate, useful as a crosslinking agent for coatings and plastics (51).

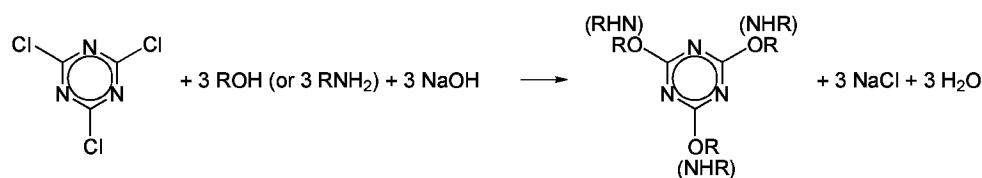
Organogermanium, organosilicon, and organotin substituted isocyanurates have been prepared by reaction of CA with the appropriate reagent containing a reactive chlorine or hydroxyl: $[(\text{C}_2\text{H}_5)_3\text{Ge}]_3(\text{NCO})_3$, $[(\text{CH}_3)_3\text{Si}]_3(\text{NCO})_3$, $[(\text{CH}_3)_3\text{SiNH}]_3(\text{NCO})_3$, $[(\text{C}_4\text{H}_9)_3\text{Sn}]_3(\text{NCO})_3$ [752-58-9] and $[(\text{C}_2\text{H}_5)_3\text{Sn}]_3(\text{NCO})_3$ [752-74-9] (70–72).

Carbohydrazide is produced in 71% yield by refluxing a mixture of CA and hydrazine hydrate for 17 h (73). Cyanuric acid and ammonia react under pressure and catalysis at 350–400°C to produce melamine [108-78-1]. At this temperature, CA dissociates to HNCO which is the probable intermediate (74). Heating a mixture of CA and boric acid above 200°C, produces hexagonal boron nitride, useful in electronic applications (75). Heating CA with high boiling fatty acids at 250°C gives high yields of the corresponding nitriles (76). Metal cyanates are obtained by treating CA with alkali metal carbonates at 350–550°C in an atmosphere of carbon dioxide (77).

Conversion of CA into cyanuric chloride [108-77-0] $(\text{ClCN})_3$ by PCl_5 is another example of reaction at carbon (78). Cyanuric chloride as an imidoyl chloride reacts as an acid chloride, unlike chloroisocyanurates.



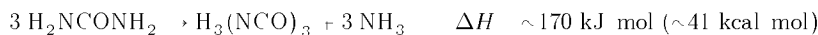
Cyanuric chloride is a convenient precursor to alkyl or aryl cyanurates by reaction with alcohols or phenols, or to substituted melamines by reaction with amines; alkaline conditions are employed in both cases and yields are generally high.



Cyanuric chloride reacts with sodium sulfide to form trithio(iso)cyanuric acid [638-16-4] (11). For a review of cyanuric chloride chemistry, see Reference 79.

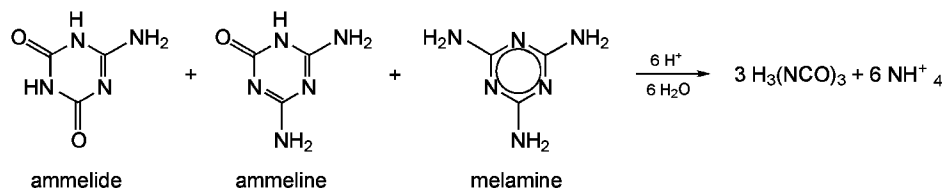
Preparation

A convenient laboratory synthesis of high purity CA is hydrolysis of cyanuric chloride (7). On a commercial scale, CA is produced by pyrolysis of urea [57-13-6]. When urea is heated at $\sim 250^\circ\text{C}$ for about an hour, it is converted to crude CA with evolution of ammonia.



The molten urea thickens and eventually solidifies. The initial decomposition reaction of urea is dissociation into isocyanic acid and ammonia. The latter two compounds recombine forming an equilibrium amount of ammonium cyanate. Biuret [108-19-0] ($\text{H}_2\text{NCONHCONH}_2$), mp 192°C , is an intermediate during the pyrolysis; some triuret [556-99-0] ($\text{H}_2\text{NCONHCONHCONH}_2$), mp 233°C , may also form. These intermediates form by successive reactions of urea with isocyanic acid. Urea cyanurate [69455-21-6] and possibly biuret cyanurate (80) are formed initially. Cyanuric acid can form by trimerization of HNCO and by cyclization of triuret. Other intermediates include cyanamide [420-04-2] and guanidine [113-00-8] which form by amination of isocyanic acid and urea, respectively. Other analogous compounds may also form by amination of biuret and triuret. These intermediates result in formation of aminotriazine by-products, primarily ammelide [645-93-2], some ammeline [645-92-1], and minor amounts of melamine [108-78-1]. Since water is formed on amination of urea and its intermediates, this results in hydrolysis of urea to NH_3 and CO_2 . Mechanisms for urea pyrolysis have been proposed (81–83).

The crude product containing aminotriazines can be purified by digestion with acids (eg, hydrochloric, nitric, or sulfuric); this hydrolyzes the acyclic impurities to carbon dioxide and ammonia and the aminotriazines to CA and ammonia.



Other options for the purification of CA include: dissolution in hot water, aqueous ammonia, aqueous formaldehyde, or hot dimethylformamide followed by filtration to remove most of the impurities. The CA is recoverable by cooling the aqueous solution (84), acidifying the ammonium hydroxide solution (85), or cooling the dimethylformamide solution with further precipitation of CA by addition of carbon tetrachloride (86). Sodium hydroxide addition precipitates monosodium cyanurate from the formaldehyde solution (87).

Cyanuric acid can also be prepared by pyrolysis of urea derivatives. Biuret and triuret give less aminotriazines due to reduced ammonia evolution. Urea cyanurate also provides a higher assay product.

By-product formation can also be reduced by use of a stripping gas or vacuum to facilitate removal of ammonia (88); however, sublimation of urea becomes excessive if the pressure is too low. Addition of ammonium salts (eg, Cl^- , NO_3^- , or SO_4^{2-}) (89–91), acids, or pyrolysis of preformed urea salts, eg, urea hydrochloride [506-89-8], sulfate, or nitrate [124-47-0] (92,93) significantly reduces aminotriazine formation. However, these alternatives present severe corrosion problems.

Aminotriazine formation can be reduced to an acceptable level ($<1\%$), thus eliminating the need for acid digestion, by pyrolysis of urea in certain high boiling solvents. Numerous solvents are disclosed in the patent literature. Some of these solvents (eg, hydrocarbon oils, aromatics, and chlorinated organics) do not dissolve urea and therefore still produce a crude product. Some solvents react with urea forming intermediates, eg, in alcohols, carbamates and allophanates are formed that decompose to CA. These intermediates can produce undesirable by-products, eg, olefins and organic carbonates. Since all organic solvents undergo some oxidation and discoloration yielding discolored product that requires further purification, exclusion of air is crucial. Chemical and physical solvent losses can render a process uneconomical. Desirable solvents are: good solvents for urea, poor solvents for CA, high boiling, and stable to pyrolysis intermediates, ammonia, oxygen, and heat. Although no perfect solvent has been identified, some solvents, eg, dinitriles (94), pyrrolidinones (95,96), and sulfones (97) largely meet these requirements.

In addition to pyrolysis of urea in certain high boiling solvents, high purity CA may be produced by: thermal decomposition of carbamyl chloride (or allophanic chlorides) produced by the high temperature reaction of ammonia and phosgene (98), and hydrolysis of cyanuric chloride, or acid digestion of aminotriazines, eg, melamine (99). These methods are not commercially viable due primarily to high raw materials costs and/or unwanted by-product formation. However, the conversion of by-product or waste cyanuric chloride or aminotriazines to CA on a break-even or near break-even basis may be economic as a pollution control measure.

Cyanuric acid can also be prepared from HNCO (100). Isocyanic acid [75-13-8] can be synthesized directly by oxidation of HCN over a silver catalyst (101) or by reaction of H_2 , CO, and NO (60–75% yield) over palladium or iridium catalysts at $280\text{--}450^\circ\text{C}$ (102). Ammonium cyanate and urea are by-products of the latter reaction.

Manufacture

The majority of the cyanuric acid produced commercially is made via pyrolysis of urea [57-13-6] (mp 135°C) primarily employing either directly or indirectly fired stainless steel rotary kilns. Small amounts of CA are produced by pyrolysis of urea in stirred batch or continuous reactors, over molten tin, or in sulfolane. The feed to the kilns can be either urea solid, melt, or aqueous solution. Since conversion of urea to CA is endothermic and goes through a plastic stage, heat and mass transport are important process considerations. The kiln operates under slight vacuum. Air is drawn into the kiln to avoid explosive concentrations of ammonia (15–27 mol %).

Concentrations of urea, CA, intermediates and by-products vary along the length of the kiln. In the later stages of pyrolysis heat transfer becomes poor and care must be exercised to prevent the reaction mass from being heated above 300°C since yields are lowered by decomposition of the product. The crude product, consisting of 2–20 mm pebbles, typically contains $\sim 20\%$ aminotriazines. Another problem occurring during pyrolysis is adhesion of the reaction products to the wall of the kiln. The resultant crude CA contains a higher level (20–30%) of aminotriazines. The gases from the kiln consisting of air, ammonia, CO_2 , urea, CA, and probably HNCO and cyanamide are scrubbed in aqueous urea and the resultant solution containing small amounts of biuret, CA, and aminotriazines is recycled to the urea feed to the kiln.

Purified CA ($>99\%$ assay) is produced by digesting the crude product in 15–20% sulfuric acid. The digestion can be carried out at atmospheric pressure or above (103–105). The digested CA slurry is filtered and washed to remove residual sulfuric acid. A small portion of the filter cake ($\sim 10\%$) is dried conventionally at temperatures up to 200°C , granulated, and packaged for merchant sales. The remainder of the CA filter cake, which contains a small

amount of water, is converted to *N*-chloroisocyanurates. The ammonia by-product from digestion is typically trapped as NH₄HSO₄; this can represent a significant water pollution burden. Indeed, many patents dealing with the manufacture of CA seek alternatives to aminotriazine formation and subsequent hydrolysis. When nitric acid is employed in crude CA purification, NH₄NO₃ is formed which can be sold as fertilizer, often blended with urea. Gaseous ammonia by-product from the pyrolysis of urea can be recovered or burned and, thus, does not represent an exceptional pollution problem. Other pollution constraints associated with isocyanurate manufacture appear in the production of chloroisocyanurates. Biocidally active chlorine in waste streams must be destroyed by hydrogen peroxide or other suitable reagents, and triazine compounds must be reduced in the effluent since they represent a nitrogen pollution load. This is generally achieved by dechlorination and acidification to yield CA which can be filtered off as is or after conversion to the monosodium salt and recycled (104).

Operability (ie, pellet formation and avoidance of agglomeration and adhesion) during kiln pyrolysis of urea can be improved by low heat rates and peripheral speeds (105), sufficiently high wall temperatures (105,106), radiant heating (107), multiple urea injection ports (106), use of heat transfer fluids (106), recycling 60–90° of the crude CA to the urea feed to the kilns (105), and prior formation of urea cyanurate (108).

Beside continuous horizontal kilns, numerous other methods for dry pyrolysis of urea have been described, eg, use of stirred batch or continuous reactors, ribbon mixers, ball mills, etc (109), heated metal surfaces such as moving belts, screws, rotating drums, etc (110), molten tin or its alloys (111), dielectric heating (112), and fluidized beds (with performed urea cyanurate) (113). All of these modifications yield impure CA.

In pyrolysis employing molten tin, a flow of the urea on the surface is eventually converted to a sheet of crude CA 15–20 mm thick. After reaching the edge of the tin bath, the moving sheet falls into a mill. The resultant powdered crude CA (contaminated with tin metal) is subjected to acid hydrolysis to convert aminotriazines (30–40° o) to CA. Tin losses can amount to 15 kg t product.

Of the many solvents disclosed, only sulfolane (ie, tetramethylene sulfone) (97) and *N*-cyclohexyl-2-pyrrolidinone (96) have been employed commercially. The pyrolysis is usually carried out under reduced pressure, ie, 27–40 kPa (200–300 torr). The success of solvent pyrolysis (and urea salt pyrolysis) in reducing aminotriazine formation is due to efficient physical (or chemical) removal of ammonia. The solvent process under vacuum and in combination with wiped-film evaporation–purification offers significant processing advantages. The reactor produces a mixture of CA containing small amounts of urea, biuret, and aminotriazines; urea and biuret are volatile and codistill with the solvent for recycling.

Economic Aspects

Monsanto and Olin, in the U.S., are the world's largest bulk suppliers of cyanuric acid and chloroisocyanurates. They have a combined CA capacity of ~38,000 t y, 90° of which is converted into chlorinated derivatives. The main foreign producers of CA are Shikoku Kasei Kogyo Co., Nissan Chemical Industries Ltd., and Nippon Soda in Japan and Atochem in France. Other smaller producers of CA include: Delsa (Spain), Sigma (Italy), and Ferrel International Co., Ltd., Kuang Ming Enterprise Co., Ltd, and Tai-Yu Chemical Industries Co., Ltd. in Taiwan. There are also a number of small producers in China and in Korea. The world production of chloroisocyanurates in 1987 was estimated to be 80,700 t (U.S. 50,000 t, Japan 17,200 t, Western Europe 13,500 t) (114). Swimming pool sales represent 76° of the total U.S. consumption. During the nine year period 1978–1987, U.S. detergent and swimming pool markets for chloroisocyanurates grew at 5 and 10° o y, respectively.

Specifications and Analysis

Cyanuric acid is sold mainly in coarse granular form, >85% 2–0.15 mm (10 to 100 mesh). It is also available in powdered form. Typical analysis of commercial CA is CA >98.5%; ammelide <1%; water <0.6%; pH of 1° o slurry >2.8. For the chlorinated derivatives:

	TCCA	SDCC-H
available Cl ₂ , ° o	90–91	55–56
moisture and volatiles, ° o	0.05–0.15	12–14
pH (1° o solution)	2.7–2.9	6.0–6.5

Analysis. Cyanuric acid can be titrated as a monobasic acid in aqueous media (end point, pH ~8.5). However, this method is of limited utility because of interferences by weak acids (eg, acetic acid) and aminotriazine impurities such as ammelide. Cyanuric acid dissolved in DMSO–benzene can be titrated potentiometrically with tetrabutylammonium hydroxide in benzene–methanol (115). Although ammelide and ammeline do not interfere, melamine does; biuret interferes slightly. Since crude CA contains only small amounts of biuret and melamine, the error is small. Cyanuric acid in wastewater can be determined by amperometric titration with Hg(NO₃)₂ (116). Aqueous solutions can be analyzed by precipitation of melamine cyanurate and quantitated gravimetrically (117), volumetrically (118), or turbidimetrically (119). Since ammelide is acidic it may also interfere in this method. Aqueous solutions of CA can also be analyzed by hplc (120) and via differential pulse polarography (121). An isocyanurate specific electrode has been used to determine CA in swimming pool water (122). Solutions of CA and chloroisocyanurates can be analyzed via uv spectrophotometry (10,123). Ammelide and ammeline can be determined by uv spectrophotometry after separation from CA via their acid solubility. A preferable method is via gc employing volatile trimethylsilyl derivatives; this method allows simultaneous determination of CA, ammelide, ammeline, and melamine (124). Solid cyanuric acid and the chloroisocyanurates can be readily identified via ir (125) or x-ray diffraction. Impurities in CA are determined as follows: ammonia, distillation-acidimetry or by means of an ion-selective electrode (124); urea, urease-acidimetry, and biuret, uv spectrophotometry. Available Cl₂ on chloroisocyanurates is determined iodometrically.

Health and Safety

Acute toxicities (LD₅₀ g kg, rats) for CA and chloroisocyanurates are: CA > 5.0, SDCC 1.67, PDCC 1.22, and TCCA 0.75 (126). A review of toxicological studies on CA and its chlorinated derivatives is given in Reference 127. These studies show that the compounds are safe for use in swimming pool and spa hot tub disinfection, sanitizing, and bleaching applications when handled and used as directed.

Most uses of chloroisocyanurates are regulated by the EPA under FIFRA. Cyanuric acid (or cyanurate) is ultimately the end product of use of chloroisocyanurates in bleaching, sanitizing, and disinfection applications. Since the *N*-chloro derivatives are biocidal, biodegradation studies have centered on the residual CA. Biodegradation occurs (128).

Cyanuric acid is stable and relatively inert. Chloroisocyanurates are also stable when dry, uncontaminated, and kept away from fire or a source of high heat. They are, however, active chlorine compounds and thus are strong oxidizing agents; care must be exercised to prevent hazardous contamination. Trichloroisocyanuric acid and the salts of DCCA retain their active chlorine during long-term storage. In addition, both TCCA and SDCC^H are resistant to sustained thermal decomposition. Upon being subjected to an intense heat source, TCCA volatilizes; SDCC^H when stored in proper containers has a greatly retarded rate of thermal decomposition in bulk as compared to the anhydrous dichloroisocyanurates. The reduced rate of thermal decomposition of the hydrate is due to hydrate bond breaking and the heat capacity and heat of vaporization of water.

Uses

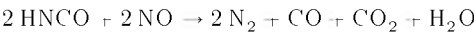
Most of the CA produced commercially is chlorinated to produce SDCC, SDCC^H, PDCC, TCCA, and the double salt TCCA4PDCC. These have become standard ingredients in formulations for scouring powders, household bleaches, institutional and industrial cleansers, automatic dishwasher compounds, and general sanitizers, and most importantly, in swimming pool and spa hot tub disinfection. The choice of chloroisocyanurate for any particular

application depends on the desired combination of solubility, av Cl₂, and pH. The salts of DCCA have a much greater water solubility than TCCA. However, TCCA has the highest av Cl₂ content and dissolves much slower than the salts. This is used to advantage in swimming pool sanitation where TCCA is tableted and dispensed into the water by means of erosion-type feeders, either floating on the surface or in-line with the filtration system. The salts are usually broadcast into the swimming pool or spa. The higher cost of the various chloroisocyanurates over less expensive chlorine bleaches such as hypochlorites is offset by better storage stability, ease of handling, and dispensing, solubility, and formulation stability with adjunct ingredients (see BLEACHING AGENTS).

Cyanuric acid is widely used to stabilize av Cl₂ against decomposition by sunlight (ie, ultraviolet light) in swimming pools; near-optimal stabilization is achieved with 25–50 ppm CA (129). Since CA is the by-product of chloroisocyanurate usage, av Cl₂ stabilization is most efficiently effected by combining an initial CA predose with routine maintenance dosing with chloroisocyanurate. An effective steady state CA level is maintained due to a balance between CA addition and losses due to filter backwashing and swimmer splashout. Recommendations for using CA and chloroisocyanurates in swimming pool disinfection can be found in Ref. 130 (see WATER, TREATMENT OF SWIMMING POOLS).

The chloroisocyanurates can be used in the bleaching of cotton, synthetics, and their blends; they do, however, attack proteinaceous fibers, such as silk or wool, presumably via active chlorine reaction with the peptide (amide) linkage. However, the chloroisocyanurates can be used as shrink-proofing agents in wool finishing (131), (see TEXTILES; WOOL). The same action of chlorine upon proteins contributes to the effectiveness of chloroisocyanurates in automatic dishwashers.

Cyanuric acid is used on a small scale for reducing nitrogen oxides (NO_x) in stationary diesel engine exhaust gases. It can also be used for NO_x reduction in coal, oil, or gas fired boilers. The CA is heated to sublime it into the exhaust gases where it dissociates to HNCO. The HNCO rapidly reacts with nitrogen oxides, via initial postulated cleavage into NH and CO radicals, forming nitrogen (132). The reaction with NO is



Melamine cyanurate is useful in preparation of flame retardant polyamide resins and compositions (133). It also is useful as a solid lubricant (134).

Tris(2-hydroxyethyl)isocyanurate (THEIC) is used as an additive in the production of high performance polyester magnet–wire enamels and in electrical varnishes. THEIC increases adhesion and imparts improved resistance to heat and weather in a variety of resin systems. In addition, the use of THEIC produces a significantly improved combination of mechanical, chemical, and electrical properties. The enamels and varnishes are used in the manufacture of electric motors, television tubes, and transformer coils (54). Homopolymers of triallyl cyanurate (TAC) and triallyl isocyanurate (TAIC) are brittle and find little application. However, TAC and TAIC can be used as crosslinking agents [eg, with polyethylene, poly(vinyl chloride) and ethylene–propylene systems (135)] and as curing agents. Either TAC or TAIC can be used as comonomer with unsaturated polyester resins (see POLYESTERS, UNSATURATED) or other unsaturated monomers including diallyl phthalate, methyl methacrylate, and styrene to form a variety of materials with enhanced properties. Tris(2,3-epoxypropyl)isocyanurate (ie, triglycidyl isocyanurate) is used as a cross-linking agent for carboxylated polyesters and as curing agent for weather-resistant powder coatings (qv) (136).

BIBLIOGRAPHY

"Triazinetriol" in *ECT* 2nd ed., Vol. 20, pp. 662–671, by Raymond N. Mesiah, FMC Corporation; "Cyanuric and Isocyanuric Acids" in *ECT* 3rd ed., Vol. 7, pp. 397–410, by Joseph V. Burakevich, FMC Corporation.

1. J. Elguero and co-workers, in A. R. Katritzky and A. J. Boulton, eds., *Tautomerism of Heterocycles*, Suppl. 1, Academic Press, Inc., New York, 1976, pp. 138–139.
2. M. Cignitti and L. Paoloni, *Spectrochim. Acta* **20**, 211 (1964).
3. A. R. Katritzky and J. M. Lagowski, in A. R. Katritzky, ed., *Advances in Heterocyclic Chemistry*, Vol. 1, Academic Press, Inc., New York, 1963, p. 387.
4. H. G. M. De Wit, J. C. Van Miltenburg, and C. G. De Kruif, *J. Chem. Thermodyn.* **15**, 651 (1983).
5. A. Schmidt, *Monatsh. Chem.* **101**, 866 (1970).
6. F. W. Hoover, H. B. Stevenson, and H. S. Rothrock, *J. Org. Chem.* **28**, 1825 (1963).
7. F. Zobrist and H. Schinz, *Helv. Chim. Acta* **35**, 2380 (1952).
8. W. Kern, H. Paul, and W. Mehren, *Makromol. Chem.* **14**, 146 (1954).
9. J. A. Wojtowicz, technical data, Olin Corp., Cheshire, Conn., 1985.
10. J. E. O'Brien, dissertation, Harvard University, Cambridge, Mass., 1972.
11. E. M. Smolin and L. Rapoport, in A. Weissberger, ed., *The Chemistry of Heterocyclic Compounds*, Vol. 13, Wiley-Interscience, New York, 1967, pp. 17–146.
12. R. H. Widman, *J. Chem. Phys.* **13**, 2922 (1965).
13. *Cyanuric Acid*, technical data bulletin, FMC Corp., Philadelphia, Pa.), 1965.
14. V. I. Zagranichnyi and Z. A. Polyakova, *Khim. Prom.*, 692 (1963).
15. E. Billows, *Z. Kristollogr. Mineral* **46**, 481 (1908).
16. **U.S. Pat. 4,275,202** (June 23, 1981), R. E. Hall and B. A. Guiliano (to FMC Corp.).
17. **U.S. Pat. 4,075,094** (Feb. 21, 1978), R. H. Carlson (to FMC Corp.).
18. V. D. Selivanov, V. M. Karlik, and V. I. Zagranichnyi, *Zh. Fiz. Khim.* **47**, 476 (1973).
19. M. K. Hargreaves, J. G. Pritchard, and H. R. Dave, *Chem. Rev.* **70**, 439 (1970).
20. **Ger. Pat. 3,224,766** (Jan. 5, 1984), H. Harth, M. Witthaus, and W. Gress (to Henkle KGaA).
21. D. Matte and co-workers, *Can. J. Chem.* **68**, 307 (1990).
22. **U.S. Pat. 2,969,360** (Jan. 24, 1961), R. H. Westfall (to FMC Corp.).
23. **U.S. Pat. 3,668,204** (June 6, 1972), R. N. Mesiah (to FMC Corp.); **U.S. Pat. 3,712,891** (Jan 23, 1973), S. Berkowitz and R. N. Mesiah (to FMC Corp.).
24. **U.S. Pat. 3,993,649** (Nov. 23, 1976), D. L. Sawhill and H. W. Schiessl (to Olin Corp.).
25. **U.S. Pat. 4,386,202** (May 31, 1983), G. D. Nelson (to Monsanto Co.).
26. **U.S. Pat. 4,122,268** (Oct. 24, 1978), S. Berkowitz (to FMC Corp.).
27. **U.S. Pats. 2,184,883; 2,184,886; and 2,184,888** (Dec. 26, 1939), I. E. Muskat and A. G. Chenicek (to Pittsburgh Plate Glass Co.).
28. **U.S. Pat. 4,007,182** (Feb. 8, 1977), J. A. Wojtowicz (to Olin Corp.); **U.S. Pat. 4,055,719** (Oct. 25, 1977), J. A. Wojtowicz (to Olin Corp.).
29. **U.S. Pat. 2,913,460** (Nov. 7, 1959), A. G. Brown, W. W. Lee, and R. M. Sancier (to The Procter & Gamble Co.).
30. **U.S. Pat. 3,894,018** (July 8, 1985), J. A. Wojtowicz and W. C. Herms (to Olin Corp.).
31. **U.S. Pat. 3,299,060** (Jan. 17, 1967), S. J. Kovalsky and R. A. Olson (to FMC Corp.) and **3,035,054** (May 15, 1962), W. F. Symes and N. S. Hadzeczyriakides (to Monsanto Co.).
32. **U.S. Pat. 3,803,144** (Apr. 9, 1974) and **3,818,004** (June 18, 1974), S. Berkowitz (to FMC Corp.).
33. **U.S. Pat. 3,035,057** (May 15, 1962), W. F. Symes and N. S. Hadzeczyriakides (to Monsanto Co.).
34. **U.S. Pat. 3,894,017** (July 8, 1975), J. A. Wojtowicz and Wade Wolfe, Jr. (to Olin Corp.); **U.S. Pat. 3,898,223** (Aug. 5, 1975) and **3,888,856** (June 10, 1975), J. A. Wojtowicz (to Olin Corp.).
35. **U.S. Pat. 4,599,411** (July 8, 1986), J. A. Wojtowicz (to Olin Corp.).
36. **U.S. Pat. 4,273,928** (June 16, 1981), S. Berkowitz (to FMC Corp.).

37. U.S. Pat. **4,024,140** (May 17, 1977) and **4,090,023** (May 16, 1978), J. A. Wojtowicz (to Olin Corp.).

38. U.S. Pat. **3,988,336** (Oct. 26, 1976), J. A. Wojtowicz (to Olin Corp.).

39. U.S. Pat. **3,272,813** (Sept. 13, 1966), W. F. Symes (to Monsanto Co.).

40. U.S. Pat. **4,103,088** (July 25, 1978), S. Berkowitz (to FMC Corp.).

41. U.S. Pat. **3,055,889** (Sept. 25, 1962) and **3,115,493** (Dec. 24, 1963), R. Marek (to Olin Corp.); U.S. Pat. **3,294,690** (Dec. 27, 1966), E. A. Matzner (to Monsanto Co.); U.S. Pat. **3,979,382** (Sept. 7, 1976), F. Martinez-Alvarez and A. G. Briones (to Sociedad Anonima).

42. G. D. Nelson, *Swimming Pool Disinfection with Chlorinated Cyanurates*, special report 6862, Monsanto Co., St. Louis, Mo., Mar. 8, 1969.

43. U.S. Pats. **3,147,254** and **3,147,259** (Sept. 1, 1964), L. O. Paterson (to Drug Research).

44. K. Morita, *Bull. Chem. Soc. Jpn.* **31**, 347 (1958).

45. W. Gottardi, *Monatsh. Chem.* **98**, 507 and 1614 (1967); **108**, 1067 (1971).

46. J. H. Saunders and R. J. Slocombe, *Chem. Rev.* **43**, 203 (1948); R. G. Arnold, J. A. Nelson, and J. J. Verbanc, *Chem. Rev.* **57**, 47 (1957).

47. N. Milstein, *J. Chem. Eng. Data* **13**, 275 (1968).

48. R. W. Cummins, *J. Org. Chem.* **28**, 85 (1963).

49. T. C. Frazier, E. D. Little, and B. E. Lloyd, *J. Org. Chem.* **25**, 1944 (1960).

50. G. T. Esayan and co-workers, *Arm. Khim. Zh.* **38**, 300 (1985).

51. U.S. Pat. **4,326,057** (Apr. 20, 1982), S. M. Cohen and J. R. LeBlanc (to Monsanto Co.).

52. A. A. Sayigh and H. Ulrich, *J. Chem. Soc.*, 3148 (1961).

53. U.S. Pat. **3,194,810** (July 13, 1965), R. L. Formaini and E. D. Little (to Allied Chemical Corp.).

54. *Tris(2-hydroxyethyl)isocyanurate*, technical data bulletin, Allied Chemical Corp., Morristown, N.J., 1960.

55. U.S. Pat. **2,809,942** (Oct. 15, 1957), H. G. Cooke, Jr. (to Devoe and Reynolds Co., Inc.).

56. *Jpn. Kokai* **76,133,288** (Nov. 16, 1976), K. Nagano and T. Fugita (to Nissan Chemical Industries, Inc.).

57. U.S. Pat. **3,075,979** (Jan. 29, 1963), J. J. Tazuma and R. Miller (to FMC Corp.).

58. *Brit. Pat.* **988,631** (Apr. 7, 1965), (to Spencer Chemical Co.).

59. *Brit. Pat.* **961,624** (June 24, 1964), (to Spencer Chemical Co.).

60. V. G. Vodop'yanov, V. G. Golov, N. P. Dvoeglazova, and M. G. Ivanov, *Zh. Obshch Khim.* **51**, 2377 (1981).

61. S. G. Grigoryan and co-workers, *Prum-st.* **3**, 34 (1981).

62. U.S. Pat. **3,249,607** (May 3, 1966), B. Taub and J. B. Hino (to Allied Chemical Corp.).

63. K. H. Slotta and R. Tschesche, *Ber.* **60B**, 301 (1927).

64. L. Paoloni and M. L. Tosato, *Ann. Chim. (Rome)* **54**, 897 (1964).

65. Z. N. Pazenko and L. I. Chovnik, *Ukr. Khim. Zh.* **30**, 195 (1964).

66. *Eur. Pat. Appl.* **EP246,190** (Nov. 19, 1987), W. Stegmann and co-workers (to Ciba-Geigy AG).

67. U.S. Pat. **4,617,391** (Oct. 14, 1986) and **4,657,939** (Apr. 14, 1987), A. Goel (to Ashland Oil Inc.).

68. U.S. Pat. **3,318,888** (May 9, 1967), J. H. Blumbergs and D. G. MacKellar (to FMC Corp.).

69. S. Freireich, D. Gertner, and A. Zilkha, *J. Organometal. Chem.* **35**, 303 (1977).

70. U.S. Pat. **3,326,906** (June 20, 1967), W. A. Stamm (to Stauffer Chemical Co.).

71. W. R. Lusby and P. C. Kearny, *J. Agric. Food Chem.* **26**, 635 (1978).

72. A. S. Gordetsov and co-workers, *Metalloorg. Khim.* **2**, 1129 (1989).

73. U.S. Pat. **3,258,487** (June 28, 1966), C. S. Argyle (to Whiffen and Sons).

74. U.S. Pat. **3,112,312** (Nov. 26, 1963), P. L. Veltman and E. Fisher (to W. R. Grace and Co.).

75. *Brit. Pat.* **951,280** (Mar. 4, 1964), T. E. O Connor (to E. I. du Pont de Nemours & Co., Inc.); U.S. Pat. **4,784,978** (Nov. 15, 1988), T. Ogasawara and co-workers (to Kawasaki Steel Corp.).

76. U.S. Pat. **2,444,828** (July 6, 1948), W. Kaplan (to Sun Chemical Corp.).

77. *Brit. Pat.* **710,143** (June 9, 1954), B. Bann and S. A. Miller (to British Oxygen Co.).

78. Z. Yoshida and R. Oda, *J. Chem. Soc. (Jpn)* **56**, 92 (1953).

79. H.-F. Piepenbrink, in E. Möller, ed., *Methoden der Organischen Chemie (Houben-Weyl), Sauerstoffverbindungen III*, Vol. 8, Georg Thieme Verlag, Stuttgart, FRG, 1952, pp. 226 ff.

80. J. Herzig, *Monatsh.* **2**, 410 (1881).

81. N. I. Malkina and S. N. Kazarnovski, *Zh. Prikl. Khim.* **34**, 1583 (1961).

82. G. Ostrogovich and R. Bacaloglu, *Rev. Roumaine Chim.* **10**, 1125 (1965).

83. A. G. Koryakin and co-workers, *Zh. Org. Khim.* **12**, 977 (1976).

84. *Jpn. Pat.* **56,079,678** (June 30, 1981), (to Mitsui Toatsu Chemicals, Inc.).

85. U.S. Pat. **3,172,886** (Mar. 9, 1965), I. Christoffel and D. P. Schutz (to Allied Chem. Corp.).

86. U.S. Pat. **2,905,671** (Sept. 22, 1959), J. D. Christian and E. W. Lard (to W. R. Grace & Co.).

87. U.S. Pat. **2,712,002** (June 28, 1955), L. A. Lundberg (to American Cynamid Co.).

88. U.S. Pat. **3,093,641** (Dec. 12, 1968), R. L. Formaini (to Allied Chem. Corp.); N. P. Shishkin and A. I. Finkel'shtein, *J. Appl. Chem. (USSR)* **42**, 446 (1969); *Zh. Prikl. Khim.* **42**, 474 (1969).

89. U.S. Pat. **2,527,316** (Oct. 24, 1950), J. S. MacKay (to American Cynamid Co.).

90. N. I. Malkina and S. N. Kazarnovskii, *Zh. Prikl. Khim.* **37**, 1158 (1964).

91. U.S. Pat. **4,093,808** (June 6, 1978), G. D. Nelson (to Monsanto Co.).

92. F. J. Schiltknecht, thesis No. 3403, Swiss Federal Inst. of Tech., 1963; *Chem. Abstr.* **63**, 5647g (1965); N. I. Malkina, *Zh. Prikl. Khim.* **34**, 1630 (1961).

93. U.S. Pat. **4,656,271** (Apr. 7, 1987), J. A. Wojtowicz (to Olin Corp.).

94. U.S. Pat. **4,360,671** (Nov. 23, 1982), J. A. Wojtowicz and D. A. Farmer, Jr. (to Olin Corp.).

95. U.S. Pat. **4,356,303** (Oct. 26, 1982) and **4,423,216** (Dec. 27, 1983), J. A. Wojtowicz and H. Hooks (to Olin Corp.).

96. U.S. Pat. **3,635,968** (Jan. 18, 1972), Goeltz and co-workers, and **3,954,751** (May 4, 1976), H. Fuchs and co-workers (to BASF AG); U.S. Pat. **4,216,315** (Aug. 5, 1980), H. H. Lewis, N. E. Morganson, and R. J. Smialek (to Olin Corp.).

97. U.S. Pat. **3,563,987** (Feb. 16, 1971), S. Berkowitz, **4,187,376** (Feb. 5, 1980), R. H. Carlson, **4,282,259** (Aug. 4, 1981), B. A. Guiliano, H. A. Pfeffer, III, and A. D. Kurtz, and **4,294,962** (Oct. 13, 1981), E. A. Bagnall, B. A. Guiliano, and H. A. Pfeffer, III (to FMC Corp.); **4,567,258** (Jan. 28, 1986), W. B. Dodge (to Olin Corp.).

98. U.S. Pat. **2,045,111** (June 23, 1936), E. Theis (to I. G. Farbenindustrie AG); *Brit. Pat.* **570,715** (July 19, 1945), D. Brundrit (to Imperial Chemical Industries Ltd.); R. J. Slocombe, E. E. Hardy, J. H. Saunders, and R. L. Jenkins, *J. Org. Chem.* **72**, 1888 (1950).

99. U.S. Pat. **2,768,167** (Oct. 23, 1956), W. F. Marzluff and L. H. Sutherland (to American Cynamid Co.).

100. U.S. Pat. **4,359,575** (Nov. 16, 1982), R. A. Sills and D. S. Katz (to American Cynamid Co.).

101. *Brit. Pat.* **883,110** (Nov. 22, 1961), (to Rohm & Haas, GmbH).

102. R. J. H. Voorhoeve and L. E. Trimble, *Science* **202**, 525 (1978); U.S. Pat. **4,174,377** (Nov. 13, 1979), (to Bell Telephone Laboratories, Inc.).

103. U.S. Pat. **4,963,674** (Oct. 16, 1990), R. W. Mason and T. C. Parker (to Olin Corp.).

104. U.S. Pat. **3,835,136** (Sept. 10, 1974) and **3,846,424** (Nov. 5, 1974), L. Hirdler, H. Schiessl, and D. Doonan (to Olin Corp.).

105. U.S. Pat. **2,943,088** (June 28, 1960), R. H. Westfall (to FMC Corp.); *Brit. Pat.* **877,725** (Sept. 20, 1961).

106. U.S. Pat. **3,953,443** (Apr. 27, 1976), O. Yoichi and A. Masayuki (to Nissan Chemical Industries, Ltd.).

107. **Jpn. Pat. 82,112,380** (July 13, 1982), (to Nippon Soda).

108. **U.S. Pat. 3,318,887** (May 9, 1967), W. P. Moore and C. B. R. Fitz-William, Jr. (to Allied Chem. Corp.) and **3,154,545** (Oct. 27, 1964), W. F. Symes and co-workers (to Monsanto Co.).

109. **Fr. Pat. 1,336,672** (Sept. 6, 1963), R. Becanne (to Office National Industriel de l'Azote); **U.S. Pat. 3,336,309** (Aug. 15, 1967), R. L. McBrayer and N. S. Nichols (to Wyandotte Chemical Corp.); **Jpn. Pat. 27,182** (Nov. 29, 1965), S. Tadao and N. Sugao (to Toa Gosei Chemical Industry, Co., Ltd.); A. Rieche, A. Gnischel, and G. Mathauser, *Chem. Tech.* (Berlin) **10**, 20 (1958).

110. **U.S. Pat. 3,093,641** (June 11, 1963), R. L. Formaini (to Allied Chem. Corp.).

111. **U.S. Pat. 3,275,631** (Sept. 2, 1966), H. Yanagizawa (to Shikoku Kasei Kogyo Co.).

112. PCT Pat. Appl. WO 79 00566 (Aug. 23, 1979), H. Kizawa and co-workers (to Nippon Soda Co., Ltd.); **U.S. Pat. 4,645,577** (Feb. 24, 1987), L. Friedman and J. A. Hamilton (to United Chemical Corp.).

113. **U.S. Pat. 3,394,136** (July 23, 1968), W. P. Moore and D. E. Elliott (to Allied Chem. Corp.).

114. *Chlorinated Isocyanurates*, *Chemical Economics Handbook*, SRI International, Menlo Park, Calif., 1988.

115. R. Morales, *Anal. Chem.* **40**, 1148 (1968).

116. V. P. Sisarenko, B. P. Zhantalai, and V. T. Chuiko, *Zh. Anal. Khim.* **32**, 2260 (1977).

117. L. Nebbia and co-workers, *Chim. Ind.* (Milan) **139**, 81 (1957).

118. **World Pat. 8,906,355** (July 13, 1989), C. O. Rupe (to Environmental Test Systems, Inc.).

119. T. Aoyama and co-workers, *Nagoya-shi Esei Kenk.* **35**, 114 (1989).

120. P. Beilstein, A. M. Cook, and R. Huetter, *J. Agric. Food Chem.* **29**, 1132 (1981); T. Sugita and co-workers, *Bull. Environ. Contam. Toxicol.* **44**, 567 (1990); A. Ghiorghis and A. H. Talebian, *J. Liq. Chromatogr.* **14**, 2331 (1991).

121. J. Struys and P. M. Wolfs, *Anal. Chim. Acta* **199**, 173 (1987).

122. **U.S. Pat. 4,134,798** (Jan. 16, 1979), M. L. Pinsky (to FMC Corp.).

123. C. J. Downes and co-workers, *Water Research*, **18**, 277 (1984).

124. Z. A. Malumyan and co-workers, *Prom-st. Arm.*, 43 (1985).

125. C. J. Pouchert, *The Aldrich Library of Infrared Spectra*, 3rd ed., Aldrich Chemical Co., Inc., Milwaukee, Wis., 1989.

126. E. Canelli, *Am. J. Public Health* **64**, 155 (1974).

127. B. G. Hammond and co-workers, *Env. Health Perspect.* **69**, 287 (1986).

128. **Brit. Pat. Appl. 2,127,006** (Apr. 4, 1984), E. A. Bagnall, M. M. Gurvitch, and R. L. Horner (to FMC Corp.); J. Saldick, *Appl. Microbiol.* **28**, 1004 (1975); D. C. Wolf and J. P. Martin, *J. Environ. Qual.* **4**, 134 (1975); H. L. Jensen and A. S. Abdel-Ghaffar, *Arch. Mikrobiol.* **67**, 1 (1969).

129. **U.S. Pat. 2,988,471** (June 13, 1961), R. J. Fuchs and I. A. Lichtman (to FMC Corp.).

130. *Pool/Spa Operators Handbook*, National Swimming Pool Foundation, San Antonio, Tex., 1988.

131. **Ger. Offen. 2,326,463** (Dec. 19, 1974), H. Bille and T. Siemenc (to BASF AG); **Jpn. Kokai 74,110,988** (Oct. 22, 1974), T. Hanyu and co-workers (to Toyobo Co., Ltd.).

132. R. A. Perry and D. L. Siebers, *Nature* **324**, 657 (1986); J. A. Caton and D. L. Siebers, *J. Eng. Gas Turbines Power* **111**, 387 (1989); J. A. Caton and D. L. Siebers, *Comb. Sci. Technol.* **65**, 277 (1989), *Comb. Flame* **79**, 31 (1990); R. K. Lyon and J. A. Cole, *Comb. Flame* **82**, 435 (1990); **U.S. Pat. 4,731,231** (Mar. 15, 1988), **4,886,650** (Dec. 12, 1989), and **4,908,193** (Mar. 13, 1990), R. A. Perry; M. P. Heap and co-workers, *Nature* **335**, 620 (1988).

133. **U.S. Pat. 4,298,518** (Nov. 3, 1981), O. Yasuhiro, M. Yukinobu, and H. Rioji (to Mitsubishi Electric Industries, Ltd.); **U.S. Pat. 4,360,616** (Nov. 13, 1982), R. U. Pagilagan (to E. I. du Pont de Nemours & Co., Inc.); **U.S. Pat. 4,504,611** (Mar. 12, 1985), C. S. Ilardo and J. J. Duffy (to Occidental Chemical Corp.).

134. T. Wang and D. Zhang, *Guti Rumbua* **9**, 137 (1989); **Jpn. Pat. 02,016,193** (Jan. 19, 1990), Y. Nakanishi and co-workers (to Kyodo Fat and Oil Co., Ltd.).

135. C. E. Schildknecht, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, Vol. 4, John Wiley & Sons, Inc., New York, 1986, pp. 802–811; *Triallyl Isocyanurate*, product bulletin, Allied Chemical Corp., Morristown, N.J., 1973.

136. *Chem. Mark. Rep.* **5** (June 20, 1977).

J. A. Wojtowicz
Olin Corporation

CYCLOHEXANE.

See HYDROCARBONS, CYCLOHEXANE.

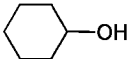
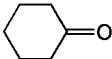
CYCLOHEXANOL AND CYCLOHEXANONE

Cyclohexanol [108-93-0] is a colorless, viscous liquid with a camphoraceous odor. It is used chiefly as a chemical intermediate, a stabilizer, and a homogenizer for various soap detergent emulsions, and as a solvent for lacquers and varnishes. Cyclohexanol was first prepared by the treatment of 4-iodocyclohexanol with zinc dust in glacial acetic acid, and later by the catalytic hydrogenation of phenol at elevated temperatures and pressures. Cyclohexanone [108-94-1] is a colorless, mobile liquid with an odor suggestive of peppermint and acetone. Cyclohexanone is used chiefly as a chemical intermediate and as a solvent for resins, lacquers, dyes, and insecticides. Cyclohexanone was first prepared by the dry distillation of calcium pimelate [19455-79-9], $\text{OOC}(\text{CH}_2)_5\text{COO}^- \text{Ca}^{2+}$, and later by Bouveault by the catalytic dehydrogenation of cyclohexanol.

Physical Properties

Important physical properties of cyclohexanol and cyclohexanone are shown in Table 1. Cyclohexanol is miscible in all proportions with most organic solvents, including those customarily used in lacquers. It dissolves many oils, waxes, gums, and resins.

Table 1. Properties of Cyclohexanol and Cyclohexanone

Property	Cyclohexanol	Cyclohexanone
		
mp, °C	25.15	47
bp, °C	161.1	156.7
d_4^{20} , g/mL	0.9493	0.9478
n_D^{20}	1.4648	
n_D^{20}		1.4507
vapor pressure, kPa ^a		
25°C	0.15	
20°C		0.53
sp heat, 15–18°C, J/g ^b	1.75	1.81
viscosity, 25°C, mPas(cP)	4.6	2.2
flash point, open cup, °C	67.2	54
solubility in water, g/100 g		
10°C	4.2	15
30°C	4.3	5
solubility of water in compound, g/100 g, 20°C	12.6	9.5

^a To convert kPa to mm Hg, multiply by 7.5.

^b To convert J to cal, divide by 4.184.

Cyclohexanone is miscible with methanol, ethanol, acetone, benzene, *n*-hexane, nitrobenzene, diethyl ether, naphtha, xylene, ethylene glycol, isoamyl acetate, diethylamine, and most organic solvents. This ketone dissolves cellulose nitrate, acetate, and ethers, vinyl resins, raw rubber, waxes, fats, shellac, basic dyes, oils, latex, bitumen, kaure, elemi, and many other organic compounds.

Manufacture

Cyclohexanol. This alcohol is produced commercially by the catalytic air oxidation of cyclohexane or the catalytic hydrogenation of phenol. The oxidation of cyclohexane to a mixture of cyclohexanol and cyclohexanone, known as KA-oil (ketone–alcohol, cyclohexanone–cyclohexanol crude mixture), is used for most production (1). The earlier technology that used an oxidation catalyst such as cobalt naphthenate at 180–250°C at low conversions (2) has been improved. Cyclohexanol can be obtained through a boric acid-catalyzed cyclohexane oxidation at 140–180°C with up to 10% conversion (3). Unreacted cyclohexane is recycled and the product mixture is separated by vacuum distillation. The hydrogenation of phenol to a mixture of cyclohexanol and cyclohexanone is usually carried out at elevated temperatures and pressure in either the liquid (4) or in the vapor phase (5) catalyzed by nickel.

Cyclohexanone.

The ketone is produced commercially by the catalytic hydrogenation of phenol or by the catalytic air-oxidation of cyclohexane; and may be prepared by either the catalytic dehydrogenation or oxidative dehydration of cyclohexanol. The hydrogenation of phenol to cyclohexanone is the most efficient route. This process can be carried out either in the liquid phase, catalyzed by palladium on carbon (6), or in the vapor phase, catalyzed by palladium on alumina (7). Of the two processes, the liquid phase is more selective and requires much less catalyst inventory. Allied Signal Corp. operates an inherently safe liquid-phase phenol hydrogenation process at temperatures below the atmospheric boiling point of the reaction liquids, permitting over 99% selectivity at 90% conversion (8).

Vapor-phase oxidation of cyclohexane is commercially feasible, but the preferred route is liquid-phase cyclohexane oxidation (2). In the latter

process, conversions are usually below 10^o %. The product is a mixture of cyclohexanone and cyclohexanol which may be separated by distillation (2) or converted to cyclohexanone by dehydrogenation over either zinc oxide (5) or nickel or magnesium oxide (9). Mild oxidation of cyclohexanol to cyclohexanone can be carried out with (10) or without (11) catalyst.

Reactions, Derivatives, and Uses

Cyclohexanol shows most of the typical reactions of secondary alcohols. It reacts with organic acids to form esters, and with halogen acids to form the corresponding halides. Dehydrating agents convert cyclohexanol into cyclohexene [110-83-8]. Catalytic dehydrogenation or mild oxidation of cyclohexanol yields cyclohexanone. Strong oxidizing agents such as nitric acid or alkaline potassium permanganate convert cyclohexanol into adipic acid [124-04-9], HOOC(CH₂)₄COOH. Oxidation of cyclohexanol to adipic acid (qv) is the most important use; fully 90^o % of adipic acid produced in the United States is used in the manufacture of nylon-6,6, a polymer of adipic acid and hexamethylenediamine. Adipic acid may also be used to produce the hexamethylenediamine. The next important usage of cyclohexanol, pure or admixed with cyclohexanone as KA-oil, is in the production of caprolactam (qv), which is used in the manufacture of nylon-6 polymer (see POLYAMIDES).

One principal use of cyclohexanol has been in the manufacture of esters for use as plasticizers (qv), ie, cyclohexyl and dicyclohexyl phthalates. In the finishes industry, cyclohexanol is used as a solvent for lacquers, shellacs, and varnishes. Its low volatility helps to improve secondary flow and to prevent blushing. It also improves the miscibility of cellulose nitrate and resin solutions and helps maintain homogeneity during drying of lacquers. Reaction of cyclohexanol with ammonia produces cyclohexylamine [108-91-8], a corrosion inhibitor. Cyclohexanol is used as a stabilizer and homogenizer for soaps and synthetic detergent emulsions. It is used also by the textile industry as a dye solvent and kier-boiling assistant (see DYE CARRIERS).

Commercial methylcyclohexanol, CH₃C₆H₁₀OH, is a slightly viscous, straw-colored liquid which is a mixture of isomers listed in Table 2. The commercial product, made by the catalytic hydrogenation of mixed cresols, has a boiling range of 155–180°C, *d*_{15.5}²⁰, 0.924 ± 0.003 g/mL, and *n*_D²⁰, 1.461. It is used as a solvent in lacquers, as an ingredient in soap-based spot removers, as a blending agent for special textile soaps and detergents, and in the manufacture of lubricating-oil additives.

Cyclohexanone shows most of the typical reactions of aliphatic ketones. It reacts with hydroxylamine, phenylhydrazine, semicarbazide, Grignard reagents, hydrogen cyanide, sodium bisulfite, etc, to form the usual addition products, and it undergoes the various condensation reactions that are typical of ketones having α-methylene groups. Reduction converts cyclohexanone to cyclohexanol or cyclohexane, and oxidation with nitric acid converts cyclohexanone almost quantitatively to adipic acid.

The most important use of cyclohexanone is as a chemical intermediate in nylon manufacture; 97^o % of all cyclohexanone output is used either to make caprolactam for nylon-6, or adipic acid for nylon-6,6. In the caprolactam process cyclohexanone is converted to cyclohexanone oxime (mp, 89–90°C), which is rearranged with sulfuric acid to ε-caprolactam [105-60-2] (mp, 69°C). The overall efficiency is >97%. In the production of adipic acid, cyclohexanone is oxidized with nitric acid in the presence of catalysts. Cyclohexanone is also used as a solvent and thinner for lacquers, especially those containing nitrocellulose or vinyl chloride polymer and copolymers and as a general solvent for synthetic resins and polymers. Cyclohexanone is an excellent solvent for insecticides and many other similar materials. Cyclohexanone is used as a building block in the synthesis of many organic compounds, such as pharmaceuticals, insecticides, and herbicides. Cyclohexanone is used in the manufacture of magnetic and video tapes (see MAGNETIC TAPE). Cyclohexanone is not restricted under California Rule 66 and has, therefore, found use as a substitute for isophorone [78-59-1] as a solvent for resins and polymers (12) (see AIR POLLUTION).

Table 2. Boiling Points of Methylcyclohexanols

Isomer	CAS Registry Number	Boiling point, °C
2-methylcyclohexanol	[583-59-5]	164.5–165.5 ^a
3-methylcyclohexanol	[591-23-1]	173.7–174 ^b
4-methylcyclohexanol	[589-91-9]	172.5–173 ^a

^a At 99.3 kPa = 745 mm Hg.
^b At 102.5 kPa = 769 mm Hg.

Economic Aspects

Estimated annual cyclohexanone production capacities are shown in Table 3; the production is greater than 90^o % captive for caprolactam production (13). The annual cyclohexanol production is only 10 thousand metric tons. These production figures do not include KA-oil (cyclohexanol-cyclohexanone) production for adipic acid. Worldwide annual capacity for cyclohexanone is approximately 3.0 million metric tons, also primarily for caprolactam production. Projected new capacity for caprolactam could add 0.5 million metric tons worldwide in this decade.

Table 3. Estimated U.S. Annual Production Capacity of Cyclohexanone

Producer	Location	Annual capacity, ^a 10 ³ t
Allied Signal, Fibers Division	Hopewell, Virginia	330
BASF Corp., Chemicals Division	Freeport, Texas	180
DSM Chemicals North America, Inc.	Augusta, Georgia	155

^a 1992.

Specifications and Analysis

Commercial cyclohexanol is offered in two grades, ie, technical and high grade. Typical specifications are as follows (14).

	Technical	High Grade
mp, °C min	10	18
distillation range	10–19 ^o % within 1.5°C of 161.7°C	100 ^o % 159–163°C, 5–95 ^o % within 1.5°C of 161.7°C
<i>d</i> _{15.5} ²⁰ , g/mL	0.9375–0.9405	0.9425–0.9455
phenol content	none to trace	none
<i>n</i> _D ²⁰	1.4575–1.4605	1.463–1.465

Some cyclohexanol is shipped with 2.25° of methanol [67-56-1] as antifreeze.

Cyclohexanol purity is best determined by gas–liquid chromatography using a DC-710 or carbowax 20M-on-chromosorb column. Impurities such as cyclohexane, benzene, cyclohexanone, and phenol do not interfere. Other analytical procedures include acylation with acetyl chloride and pyridine at elevated temperatures. Following hydrolysis, the excess acid is titrated with sodium hydroxide solution (15). Aldehydes and easily hydrolyzed esters interfere with this method. For samples containing high concentrations of acids and easily hydrolyzed esters but no aldehydes or ketones, an alternative procedure is used (16). This procedure involves treatment at elevated temperatures of the sample containing cyclohexanol with acetic acid containing boron trifluoride catalyst. The water formed in the esterification reaction is then determined by the Karl Fischer method. For analyses based on micro procedures, acetic anhydride is the preferred acetylating agent (17).

Cyclohexanol can be determined colorimetrically by reaction with *p*-hydroxy-benzaldehyde in sulfuric acid (18). This method can be used in the presence of cyclohexanone and cyclohexane. Cyclohexanol and cyclohexanone both show a maximum absorbency at 535 nm but at 625 nm the absorption by cyclohexanone is negligible, whereas cyclohexanol shows appreciable absorption.

Cyclohexanol and cyclohexanone are shipped in 208-L (55-gal) drums, in tank cars, and tank trucks. DOT regulations classify both cyclohexanol and cyclohexanone as combustible liquids. Drums containing less than 416-L (110 gal) do not require hazardous material labeling. Larger quantities must be labeled "Combustible Liquid" (19). The price in 1990 (for Technical, tank trucks, delivered) (13), was \$1.87 kg for cyclohexanol and \$1.70 kg for cyclohexanone.

Commercial cyclohexanone is offered in various grades. The specifications of two typical grades are listed in Table 4 (14).

Table 4. Specifications of Two Typical Cyclohexanones

Property	Commercial grade	High purity
appearance	colorless liquid	colorless liquid
$d_{4}^{15.5}$, g mL		0.950–0.951
d_{20}^{20} , g mL	0.940–0.950	
distillation range, 95° at 101.3 kPa ^a	152–157°C	152–157°C
ash, ° max	0.01	0.01
acidity, ° max	none	0.03
alkalinity to phenolphthalein	none	none
ketone content, ° min	89	99.5
soly in water at 20°C, °		5
water content, ° max	0.2	0.2
flash point		
(Abel closed cup), °C		42
(Cleveland open cup), °C	46	
evaporation rate (ether = 1)	38	40

^a To convert kPa to mm Hg, multiply by 7.5

Cyclohexanone purity is most readily determined by gas-liquid chromatography over DC-710 or carbowax 20M-on-chromosorb. Impurities such as cyclohexane, benzene, cyclohexanol, and phenol do not interfere. In the absence of other carbonyl compounds cyclohexanone may be determined by treatment with hydroxylamine hydrochloride, which forms the oxime, as follows:



With materials containing low concentrations of acids, the liberated hydrogen chloride is titrated with sodium hydroxide, either electrometrically (20) or with an indicator (21).

Health and Safety Factors

Cyclohexanol is slightly toxic by the oral route of exposure and is slightly irritating to the skin (22). It can cause severe eye irritation and transient corneal injury (23). The vapor is irritating to the eyes of humans at 100 ppm (24). Inhalation of high concentrations (~1000 ppm) of cyclohexanol in rabbits for 6 h day, 5 days wk for about 10 weeks induced effects including conjunctival congestion and irritation, lacrimation, salivation, lethargy, incoordination, narcosis, and mild convulsions.

After pregnant female mice were fed a diet containing 1° cyclohexanol during gestation, the offspring were observed to have an increased mortality during the first 21 days of life (25). Since only one dose (1°) was employed, a no-observable-effect level (NOEL) could not be determined from this study. The ability of cyclohexanol to cause testicular atrophy resulting in loss of spermatogonia, spermatocytes, and spermatozoa with shrinkage of seminiferous tubules and Leydig cells was reported upon repeated subcutaneous administration of 15 mg kg day in rats (for 37 days) and gerbils (for 21 days) (26). Similar signs were observed in rabbits after repeated oral administration of 25 mg kg day for a period of 40 days. Normal spermatogenesis returned after a 70 day recovery period (27). However, in one other reported study, testicular atrophy was not observed in rats administered cyclohexanol at a dose of 455 mg kg day by daily gastric intubation for 7 days (28).

The ACGIH threshold limit value (TLV), time-weighted average for an 8 hour workday, 40 hour workweek, was set at 50 ppm (~200 mg m³) with a notation for skin absorption (29).

Precautions that should be observed as a matter of course in using cyclohexanol include adequate and proper ventilation, avoidance of prolonged breathing of vapor or contact of the liquid with the skin, avoidance of ingestion, and protection of the eyes against splashing liquids.

Cyclohexanone has only slight toxicity by the oral, dermal, and inhalation routes of exposure. Liquid or vapor exposures may result in transient corneal injury (23). Primary irritation and defatting of the skin can result from substantial or prolonged contact with cyclohexanone. Exposure to high vapor concentrations can cause central nervous system (CNS) depression, an effect which can also occur after ingestion or repeated dermal exposure. Both positive and negative results have been reported for the genetic activity of cyclohexanone (30). The preponderance of the negative data stems from *in vivo* assay systems with positive results obtained with *in vitro* assay systems. Convincing evidence for carcinogenicity has not been provided, however (30,31). Cyclohexanone has been reported to be nonteratogenic in both rats and mice (32,33). Results of a two-generation inhalation reproduction study in rats indicate that the growth, development, or reproductive performance of animals exposed to 1000 ppm for one generation, or exposed to 250 or 500 ppm for two generations, was not adversely affected. When the progeny derived from parental animals (previously exposed to 100 ppm cyclohexanone) were exposed to 1400 ppm cyclohexanone for one generation, effects observed included lethargy, male body weight depressions, reduced male fertility, reduced progeny survival, and progeny body weight depression. These effects were found to be reversible during a post-exposure recovery period. Furthermore, there was no evidence of a dominant lethal effect resulting from the exposures. A no-observable-effect level (NOEL) of at least 500 ppm cyclohexanone vapor was determined from this study (34).

The time-weighted average OSHA permissible exposure limit (PEL), as well as the ACGIH threshold limit value (TLV), for cyclohexanone is 25

ppm (100 mg m³) with a notation for skin absorption (29,35).
The precautions usually observed when handling volatile solvents should be observed as a matter of course with cyclohexanone. These include adequate and proper ventilation, avoidance of prolonged breathing of vapor or contact of the liquid with the skin, avoidance of internal consumption, and protection of the eyes against splashing liquids.

BIBLIOGRAPHY

" Cyclohexanol and Cyclohexanone" in *ECT* 1st ed., Vol. 4, pp. 769–773, by J. D. Young, E. I. du Pont de Nemours & Co., Inc.; in *ECT* 2nd ed., Vol. 6, pp. 683–688, by R. D. Kralovec and H. B. Louderback, E. I. du Pont de Nemours & Co., Inc.; in *ECT* 3rd ed., Vol. 7, pp. 410–416, by W. B. Fisher and J. F. VanPeppen, Allied Signal Inc.

1. *Chem. Week* **93** (20), 74 (1963).
2. **U.S. Pats. 2,223,493** and **2,223,494** (Dec. 3, 1940), D. J. Doder (to E. I. du Pont de Nemours & Co., Inc.).
3. "Cyclohexanol and Cyclohexanone: Salient Statistics," *Chemical Economics Handbook*, SRI International, Menlo Park, Calif., June 1976.
4. **Brit. Pat. 22,523** (Oct. 6, 1913), A. Brochet.
5. **U.S. Pat. 3,998,884** (Dec. 21, 1976), C. A. Gibson (to Union Carbide).
6. **U.S. Pat. 3,076,810** (Feb. 5, 1963), R. J. Duggan and co-workers (to Allied Chemical Corp.).
7. **U.S. Pat. 3,305,586** (Feb. 21, 1967), P. Phielix (to Stamicarbon).
8. **U.S. Pat. 4,092,360** (May 30, 1978), J. F. VanPeppen and W. B. Fisher (to Allied Chemical Corp.).
9. **USSR Pat. 232,231** (July 25, 1977) (to Scientific-Research Physical-Chemical Institute).
10. **U.S. Pat. 3,974,221** (Aug. 10, 1976), R. J. Duggan (to Allied Chemical Corp.).
11. **U.S. Pat. 2,285,914** (June 9, 1942), O. Drossbach (to E. I. du Pont de Nemours & Co., Inc.).
12. W. L. Faith, D. B. Keyes, and R. L. Clark, *Industrial Chemicals*, John Wiley & Sons, Inc., New York, 1975.
13. *Chemical Economics Handbook*, SRI International, Menlo Park, Calif., Sept. 1990; Allied Signal Co., internal data, 1992.
14. H. C. Reis and R. G. Muller, *Caprolactam*, Part II, Report No. 7, Process Economics Program, SRI International, Menlo Park, Calif., 1965, pp. 441–442.
15. D. M. Smith and W. M. D. Bryant, *J. Am. Chem. Soc.* **57**, 61 (1935).
16. W. M. D. Bryant, J. Mitchell, Jr., and D. M. Smith, *J. Am. Chem. Soc.* **62**, 1 (1940).
17. N. D. Cheronis and T. S. Ma, *Organic Functional Group Analysis by Micro and Semi-micro Methods*, Interscience Publishers, Inc., New York, 1964, pp. 493–496.
18. S. D. Nogare and J. Mitchell, Jr., *Anal. Chem.* **25**, 1376 (1953).
19. R. M. Graziano, *Hazardous Material Regulations of the Department of Transportation*, Tariff #31, Association of American Railroads, Bureau of Explosives, Washington, D.C., Mar. 31, 1977.
20. J. Mitchell, Jr., D. M. Smith, and W. M. D. Bryant, *J. Am. Chem. Soc.* **63**, 573 (1941).
21. W. M. D. Bryant and D. M. Smith, *J. Am. Chem. Soc.* **57**, 57 (1935).
22. R. W. Reinhold, Allied Signal, private communication, 1992.
23. G. D. Clayton and F. E. Clayton, eds., *Patty's Industrial Hygiene and Toxicology*, 3rd rev. ed. Vol. IIC *Toxicology*, John Wiley & Sons, Inc., New York, 1982.
24. W. Morton Grant, *Toxicology of the Eye*, 3rd ed., Charles C Thomas, Springfield, Ill., 1986.
25. E. Gondry, *J. Eur. Toxicol.* **5**(4), 227–238 (1973).
26. A. Tyagi, B. C. Joshi, S. Kumar, and V. P. Dixit, *Indian J. Expt'l Biol.* **17**, 1305–1307 (1979).
27. V. P. Dixit, R. S. Gupta, S. Kumar, and B. C. Joshi, *Indian J. Physiol. Pharmacol.* **24**(4), 278–286 (1980).
28. B. G. Lake and co-workers, *Acta Pharmacol. Toxicol.* **51**, 217–226 (1982).
29. *1991–1992 Threshold Limit Values for Chemical Substances and Physical Agents and Biological Exposure Indices*, American Conference of Government Industrial Hygienists, Cincinnati, Ohio, 1991.
30. *Integrated Risk Information System (IRIS) Database*, U.S. Environmental Protection Agency, Washington, D.C., 1992.
31. *Toxicity Profile—Cyclohexanone*, BIBRA Toxicology International, The British Industrial Biological Research Association, Carshalton, Surrey, UK, 1987.
32. *An Inhalation Teratology Study in Rats with Cyclohexanone*, a study performed by Bio dynamics, Inc., East Millstone, N.J., for The Industrial Health Foundation, Aug. 1984.
33. *An Inhalation Teratogenicity Study in the Mouse with Cyclohexanone*, a study performed by Bio dynamics, Inc., East Millstone, N.J., for The Industrial Health Foundation, Aug. 1984.
34. *Two Generation Reproduction Study via Inhalation (With A Neurotoxicology/Pathology Component) in Albino Rats Using Cyclohexanone*, a study performed by American Biogenics Corp., Decatur, Ill., for The Industrial Health Foundation, Mar. 1986.
35. *Registry of Toxic Effects of Chemical Substances (RTECS)*, NIOSH Database, Cincinnati, Ohio, 1992.

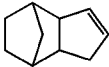
William B. Fisher
Jan F. VanPeppen
Allied Signal Inc.

CYCLOPENTADIENE AND DICYCLOPENTADIENE

Cyclopentadiene [542-92-7] (CPD), C₅H₆ (1) and its more stable and available dimer form, dicyclopentadiene [77-73-6] (DCPD), C₁₀H₁₂ (2) are two well-established and versatile chemical building blocks.



(1)



(2)

The monomer, CPD, obtained via cracking of the dimer, DCPD, and the dimer both have extensive uses. Cyclopentadiene is probably the most widely studied conjugated, cyclic diolefin system. Eleven review articles dealing with the chemistry of cyclopentadiene have been published (1–11). An article dealing specifically with European uses of DCPD has also been published (12). The discovery in 1951 of stable metal derivatives has given additional impetus to the study of the chemistry of cyclopentadiene. Five review articles have been published on this subject (13–17).

Physical Properties

Dicyclopentadiene exists in two stereoisomeric forms, the endo and exo isomers. Commercial DCPD, 3a,4,7,7a-tetrahydro-4,7-methano-1H-indene, is predominantly the endo isomer (exo:endo 6:953 by capillary gas chromatography). The dimer is the form in which CPD is sold commercially. The physical properties of CPD and DCPD are given in Table 1. The strong absorption bands in the ir and uv spectra as well as the Raman spectra are discussed in Reference 2. The nmr spectrum of DCPD is described in Reference 18.

Table 1. Physical Properties of Cyclopentadiene and Dicyclopentadiene

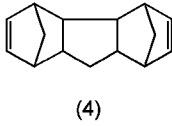
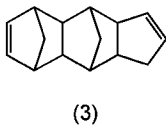
Physical properties	Cyclopentadiene	Dicyclopentadiene ^a
mol wt	66.1	132.2
bp, 101.3 kPa ^b , °C	41.5	170 ^c
mp, °C	85	33.6
physical form	colorless liquid	colorless crystals
odor	sweet terpenic	camphoraceous
d_4^{20} , g mL	0.8024	
d_4^{25} , g mL		0.9770
n_D^{20}	1.4429	
n_D^{25}		1.5061
heat of combustion, kJ mol ^d	2929	5767
heat of vaporization, kJ mol ^d	29.3	38.5
heat of cracking, kJ mol ^d		102.9
heat of fusion, kJ mol ^d		2.1
specific heat, kJ (kgK) ^d		1.7
spontaneous ignition temp, °C		510
in oxygen		680
in air	640	
dielectric constant at 40°C	2.43	

^a Endo isomer.
^b To convert kPa to mm Hg, multiply by 7.5.
^c Depolymerizes at boiling point to form two molecules of cyclopentadiene.
^d To convert kJ to kcal, divide by 4.184.

Chemical Reactions

Cyclopentadiene contains conjugated double bonds and an active methylene group and can thus undergo a Diels-Alder diene addition reaction with almost any unsaturated compound, eg, olefins, acetylene, maleic anhydride, etc. The number of its derivatives is extensive; only the reactions and derivatives considered most important are discussed.

Polymerization. CPD dimerizes spontaneously and exothermically at ambient temperature to DCPD. At temperatures above 100°C, CPD can be made to polymerize noncatalytically via a series of consecutive Diels-Alder reactions to trimer, tetramers, and higher oligomers. For example, the trimers, 3a,4,4a,5,8,8a,9,9a-octahydro-4,9:5,8-dimethanobenz-1H-[f]indene, [7158-25-0] (3) and 1,4,4a,4b,5,8,8a,9a-octahydro-1,4:5,8-dimethano-1H-fluorene [35184-08-8] (4), are formed in the ratio 87:13 by the monomer adding to the dimer (19).



The lower molecular-weight thermal oligomers are crystalline compounds with little odor. The yields of oligomers at varying contact times and different temperatures are shown in Table 2.

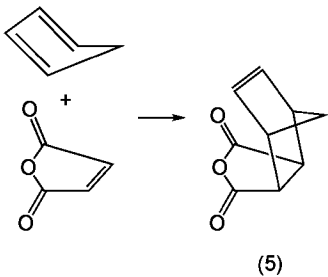
Table 2. Yield of Oligomers at Various Contact Times and Temperatures, wt %

Oligomer	At 150–160°C, 14 h	At 170–180°C, 22 h	At 200°C, 90 h
tricyclopentadiene	40	50	25
tetracyclopentadiene	10	30	45
pentacyclopentadiene	2	5	10
DCPD (unchanged)	50	10	5

Cyclopentadiene oligomers up to octamers can be effectively analyzed and quantified by supercritical fluid chromatography using a chemically bonded methyl silicone capillary column.

In addition to thermal polymerization, it is possible to polymerize CPD with inorganic halides as catalyst. With trichloroacetic acid as the catalyst, deeply colored, blue polymers that conduct electricity in nonpolar solvents such as benzene in the presence of acid can be obtained. The conductivity and color are caused by blocks of conjugated double bonds present in the polymers (20–21).

Diene Addition (Diels-Alder). The conjugated double bonds of CPD can undergo a diene addition reaction. The reaction involves the addition of an ethylenic group contained in the first reactant, called the dienophile, across the 1,4-position of cyclopentadiene. The Diels-Alder addition products are usually bicyclo[2.2.1]heptene derivatives. The reaction is carried out by simply bringing the two reactants together in the presence or absence of a solvent at temperatures ranging from nominally room temperature in some cases to about 200°C. The reaction is highly exothermic (71–75 kJ mol or 17–18 kcal mol), and in many cases the yield is essentially quantitative, especially with strong dienophiles such as acrylic acid or maleic anhydride. Although CPD monomer is usually used, it is possible to start with DCPD, provided the reaction is carried out at temperatures (usually 175–200°C) at which the DCPD dissociates into the monomer, which then adds to the dienophile. Superambient pressures are normally used in these latter reactions. With maleic anhydride as the dienophile, the reaction leads to *endocis*-bicyclo[2.2.1]hept-5-ene-2,3-dicarboxylic anhydride [129-64-6] (5), known as nadic anhydride.

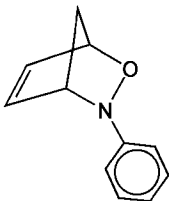
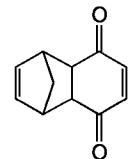
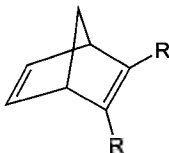
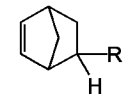
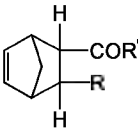
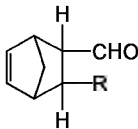


The diene addition is stereochemically specific (22). Although two stereoisomers are possible, the endo and the exo forms, in the addition of maleic anhydride to CPD the endo adduct is the exclusive product (23).

As a result of the diene addition reaction, CPD has been converted into innumerable derivatives, some of which are tabulated in Table 3. The products obtained from the diene addition reaction are extremely versatile chemicals suitable as intermediates for the production of plasticizers, pharmaceuticals, pesticides, resins, paint driers, perfumes, and many other products.

Table 3. Diels-Alder Adducts from Cyclopentadiene

Dienophile	Structure of adduct
<i>Dibasic acids and derivatives</i>	
chloromaleic anhydride, R = Cl	
maleic anhydride, R = H	
<i>Monobasic acids</i>	
crotonic acid, R = H; R' = CH ₃	
methacrylic acid, R = CH ₃ ; R' = H	
<i>Aldehydes</i>	
acrolein, R = H	
crotonaldehyde, R = CH ₃	



Ketones

methyl propenyl ketone (3-penten-2-one), R = CH₃; R' = CH₃
methyl vinyl ketone (3-buten-2-one), R = H; R' = CH₃

Vinyl compounds

ethylene, R = H
styrene, R = C₆H₅
vinyl acetate, R = OOCCH₃

Acetylenes

acetylene, R = H
acetylenedicarbonitrile, R = CN

A quinone

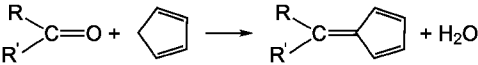
p-benzoquinone

A nitroso compound

nitrosobenzene

Condensation Involving the Methylene Group. Cyclopentadiene is a very acidic hydrocarbon, having a pK_a of 18 due to the aromatic stabilization of its six electron cyclic anion. This stabilization makes the methylene group in cyclopentadiene extremely reactive, allowing it to undergo condensation-type reactions. A large number of derivatives have been prepared this way.

Aldehydes and ketones such as acetaldehyde, benzaldehyde, acetone, acetophenone, cyclohexanone, cyclopentanone, and methyl ethyl ketone have been condensed with CPD in the presence of alkaline agents to produce colored fulvene derivatives. A typical condensation with a ketone is depicted as follows:



The color of a fulvene deepens when R and R' change from hydrogen to methyl, and deepens further with increasing conjugation with R, eg, R = vinyl. With aromatic substituents the fulvenes are deep red. The aldehyde condensation products are also strongly colored but resinify so easily that it is difficult to isolate them in the pure form.

Hydrogenation. Dicyclopentadiene has been progressively hydrogenated through the dihydro to the tetrahydro derivative (24–25). The double bond in the bicycloheptene ring is the more reactive, and the heat of hydrogenation is 139 kJ/mol (33.2 kcal/mol). The heat of hydrogenation for the five-membered ring is 110 kJ/mol (26.2 kcal/mol) (26).

The oligomers can also be hydrogenated. The trimers are easily hydrogenated, first to the dihydro and then to the tetrahydro derivative using palladium, Adams' platinum, massive nickel, or other conventional catalysts. The endo isomers of the hydrogenated compounds are waxlike solids. Dihydro-*endo*-tricyclopentadiene melts at 37°C and boils at 145–146°C at 1.7 kPa (13 mm Hg) and 271°C at 102 kPa (766 mm Hg). Tetrahydro-*endo*-tricyclopentadiene melts at 49–53°C, and boils at 151°C at 2 kPa (15 mm Hg) and ca 285°C at 102 kPa (766 mm Hg). Data on the hydrogenation process and physical properties of the hydrogenated oligomers of CPD are in References 26–28.

Halogenation. Halogens and halogen acids add readily to the unsaturated carbon linkages of the cyclopentadiene molecule. By such additions a series of halogenated derivatives range, in the case of the chloride, from 3-chlorocyclopentene to tetrachlorocyclopentane. Of all the possible chloro derivatives of CPD, only hexachlorocyclopentadiene [77-47-4] ever reached commercial status. It was used as an insecticide, but this use has been discontinued because of its toxicity (see CHLOROCARBONS AND CHLOROHYDROCARBONS, TOXIC AROMATICS). It can be prepared by a liquid phase chlorination of CPD below 50°C (29).

Oxidation. Cyclopentadiene reacts spontaneously with oxygen to form brown, gummy products that usually contain substantial amounts of

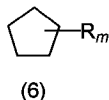
peroxides. Catalytic vapor-phase oxidation of CPD at 400–525°C over vanadium oxide leads to the production of maleic anhydride.

Dihydroxycyclopentenones and tetrahydroxycyclopentane can be prepared by the oxidation of CPD with hydrogen peroxide (30–33).

Possible uses for these polyhydroxy compounds include the preparation of alkyd-type resins with polybasic acids, the formation of ester plasticizers, and the preparation of surface-active agents.

Cyclopentadiene has also been oxidized by singlet oxygen to 4,5-epoxypenten-2-al-1, cis and trans isomers. These compounds and their hydrogenated diol products are claimed as useful intermediates as cross-linking agents, and in the production of pesticides and perfumes (34).

Alkylation. CPD can be multiply alkylated in high yields using alkyl halides, oxo alcohols, and Guerbet alcohols (35–36). After hydrogenation of the diene products, these so-called multiply alkylated cyclopentanes (6), where $R = C_4H_9 - C_{20}H_{41}$ and $m = 2-6$, have been demonstrated to be useful as synthetic lubricants.



Source and Production

No commercial process for the sole, deliberate, synthetic production of cyclopentadiene exists. It is obtained as a by-product from thermal operations. In the carbonization of coal, the tar, light oil, and coke-oven gas contain cyclopentadiene and dicyclopentadiene in very low concentrations (see COAL, CARBONIZATION). Steam-cracking (thermal cracking in the presence of steam) of hydrocarbons, ethane, propane, and particularly gas oil and naphtha, represents the principal source of CPD and also methylcyclopentadienes (MCPD) consisting of a mixture of 1-methyl-, 2-methyl- and 5-methylcyclopentadiene in the ratio 45:52:3 (37). The yield of CPD is usually less than 2 wt % from ethane–propane mixtures (E–P mix), and increases with heavier, higher molecular weight steam-cracker feeds at the same steam-cracking conditions of temperature, residence time, and steam:feed ratio.

Steam-cracking operations also produce other hydrocarbons including benzene, naphthalene, isoprene, piperylene, ethylene, propylene, etc. CPD and MCPDs are always produced in steam-cracking. Cyclopentadiene is recovered from the other hydrocarbons as follows: The total cracked product is initially separated in a primary fractionation tower. Compounds boiling lower than 230°C are sent to a splitter tower, whose cut point (final boiling point of the overhead stream) is determined by the overall economics and dispositions of the steam-cracked products. Usually this is set at the boiling point of benzene. Splitter tower bottoms are usually processed into benzene, toluene, xylenes, and a wide boiling stream whose reactive components are comprised of styrene, alkylstyrenes, indene, and alkylindenes. This latter stream is called resin oils. Cycloienes (CPD and MCPD) are sometimes present as dimers and codimers in these resin oils, which are used in hydrocarbon resin syntheses.

The splitter tower overhead, consisting of C_5 -hydrocarbons and lighter components, is processed through a series of distillation towers. In the course of this process, the distillate is heated in the reboilers of the towers to a temperature of about 100°C to convert monomeric CPD to DCPD. Depending on the temperature, concentration, and total flow, the heat-soaking residence time may be 5–24 hours. The DCPD, which boils higher than the unreacted hydrocarbons of the distillate, is recovered as distillation bottoms. This is a relatively low energy operation, with few distillations required to separate the C_5 and lighter components from the higher boiling dimers. The process provides low purity DCPD streams typically having DCPD contents of 75–85 wt %. See Figure 1 for a simplified flow diagram of the process. Regardless of its source, only the dimer is available, as the monomer spontaneously reacts to form the dimer. High-purity DCPD is obtained by cracking the DCPD in the crude stream, separating the low-boiling monomer CPD by distillation, and allowing the concentrated CPD to dimerize under controlled conditions. The process of preparing CPD from DCPD (illustrated in Ref. 38) can be used with either crude or pure DCPD streams and includes detailed temperature data on the process described.

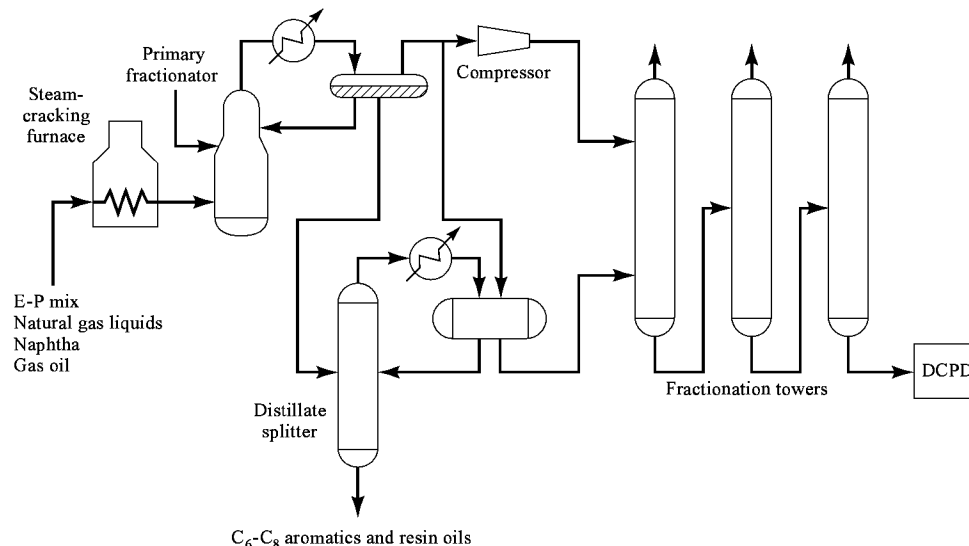


Fig. 1. Simplified DCPD recovery flow diagram. Number of towers varies among manufacturers. E–P = ethane–propane.

In the United States, Europe, and Japan, DCPD streams of 70–95 wt % purity are available. Estimates of recoverable DCPD production capacity in the United States for 1990 for all grades of DCPD is >127,000 metric tons (39) and in Europe is 48,000 metric tons (40). The vast majority of this production is from hydrocarbon steam-cracking operations. Based on the total operations, more CPD is produced than indicated above, but because of the relatively small quantities available at a single location, much of the cyclopentadiene cannot be recovered profitably. Important producers in the U.S. are Dow, Exxon, Lyondell, Shell, and Texmark; in Europe, Dow and Shell; and in Japan, Nippon Zeon (40).

The purity of a dicyclopentadiene stream may be expressed in terms of DCPD itself or in terms of available CPD monomer. Both analyses are determined by gas chromatography (gc). The first analysis is capillary gc on a nonpolar column. The data from the analysis can be used to calculate the available CPD, assuming that all the DCPD and CPD codimers crack completely. In the second analysis the sample is charged to the gc equipment under temperature conditions (injection port 400°C) that cause essentially complete reaction of the dimers to monomers.

Handling and Storage

Cyclopentadiene monomer spontaneously dimerizes at room temperature. The approximate dimerization rates of liquid CPD at temperatures normally encountered in handling CPD monomer are as follows. The rate constant $k = 1.2 \times 10^6 \exp(-16.7 [RT\epsilon])$ L (mole) (2).

Temp, °C	Dimerization rate, mol % h	Temp, °C	Dimerization rate, mol % h
20	0.05	25	3.5
0	0.5	30	6
10	1	35	9
15	1.5	40	15
20	2.5		

The dimerization is highly exothermic (ca 75 kJ mol or 18 kcal mol) and, as the above rates indicate, increases rapidly with increase in temperature. Therefore, if the temperature around a closed container is not low enough to dissipate the heat of reaction through the available surface, the temperature rises, the dimerization reaction accelerates, and the pressure can increase rapidly enough to cause a rupture of the container. Because of this spontaneous dimerization and the fact that DCPD is readily reconverted to the monomer, the dimer forms a safe and convenient way of handling CPD commercially. Generally, DCPD is moved in barges, tank cars, and trucks, all of which may be equipped with heating coils to maintain a liquid, pumpable product. However, DCPD should not be left in heated, idle lines and pumps for long times, ie, several weeks, as it will slowly polymerize and plug them.

The consumer of CPD monomer must thermally crack DCPD to obtain it. The dissociation is a monomolecular reaction. The rate is expressed by the equation $k = 6 \times 10^{12} \exp(-34,000/RT)$ liters per mole per second for DCPD in the pure liquid state (2). From the values in Table 1, the energy required to crack a kilogram of DCPD to CPD is ca 780 kJ kg (ca 740 Btu kg). Although this is not very large, it illustrates the process problem facing the user relative to other diolefins that do not require a separate operation to produce the monomer. The cracking can be accomplished readily by distilling the DCPD at atmospheric pressure. As DCPD boils at 170°C, it cracks at a rate of about 36 percent per hour; by maintaining the top temperature of the fractionating column at 41–42°C, a distillate consisting of pure CPD monomer is obtained. Another procedure involves adding DCPD to some kind of hot liquid, preferably a high-boiling oil, at 250–260°C. The resulting vapors are fractionated to remove refluxing, uncracked dimer, and other entrained liquid hydrocarbons. The most practical method, from a commercial standpoint, consists of cracking the dimer to the monomer in the vapor phase with relatively short contact times at 350–400°C. Virtually complete cracking of the dimer to the monomer occurs under these conditions.

Obtaining high purity (>99%) CPD, or alternatively high purity DCPD, from lower purity (75–85%) DCPD is not readily accomplished in high yield by straight fractional distillation. The main difficulty in this purification of DCPD is the presence of codimers of C₄ and C₅ acyclic dienes with CPD, CPD–MCPD codimers, C₄ and C₅ acyclic dienes with MCPD, and the Cope rearrangement products (because of the heat soaking) of these codimers. The distillation might be possible at economical recoveries at atmospheric pressure except that DCPD and some of the codimers crack at the temperature of distillation. At high vacuum, the differences in boiling points drop to an extent that the recovery is uneconomical except for DCPD of >94% purity. Many of the codimers themselves are difficult to synthesize and isolate in pure form in order to obtain their physical properties, eg, boiling points. Estimates of the boiling points of these codimers have been derived via their gas chromatographic retention indices (41). These are listed in Table 4.

Table 4. Estimated Boiling Points of DCPD Codimer Cycloolefins^a

Cycloolefin	Bp, °C
2- <i>endo</i> -vinylnorbornene-5	137.0
2- <i>exo</i> -vinylnorbornene-5	137.8
2- <i>exo</i> -methyl-3- <i>endo</i> -vinylnorbornene-5	150.1
2- <i>endo</i> -methyl-3- <i>exo</i> -vinylnorbornene-5	150.5
2- <i>endo</i> -methyl-2- <i>exo</i> -vinylnorbornene-5	154.3
2- <i>exo</i> -methyl-2- <i>endo</i> -vinylnorbornene-5	155.2
tetrahydroindene	160.0
2- <i>exo</i> -isopropenylnorbornene-5	165.8
2- <i>endo</i> -isopropenylnorbornene-5	166.2
2- <i>endo-trans</i> -propenylnorbornene-5	166.7
2- <i>exo-cis</i> -propenylnorbornene-5	166.9
2- <i>endo-cis</i> -propenylnorbornene-5	167.9
<i>endo</i> -dicyclopentadiene	170.0
7-methyltetrahydroindene	177.9
4-methyltetrahydroindene	179.6
5-methyltetrahydroindene	179.9
6-methyltetrahydroindene	180.5

^a Ref. 41.

Vapor-phase cracking (350–400°C) produces not only CPD, but also butadiene, vinylacetylene, isoprene, and *cis*- and *trans*-1,3-pentadienes (piperylenes) via cracking of the various codimers. Although the monomers can be separated in a superfractionation tower, the total separation of pure CPD is difficult, and much work has been devoted to this end. Most methods center on maintaining the cracking temperature below 250°C to minimize cracking of the isoprene–CPD and piperylene–CPD codimers (42–44). The CPD effluent is further purified by fractionation.

The distillation and or cracking of DCPD should be carried out in an inert atmosphere to prevent the formation of peroxides. DCPD should never be distilled to dryness since there is danger of explosion if peroxides are present. Also, when cooling a piece of equipment that contains the hot bottoms from the distillation or cracking of DCPD, care should be taken to exclude air, usually accomplished by blanketing the system with nitrogen. Contacting the hot bottoms with air could lead to rapid peroxide formation with a resulting disruptive polymerization.

If pure monomer is to be used in a reaction, it must be used immediately or stored at ≤ –20°C to prevent dimerization to any appreciable extent. Chemical inhibition does not prevent dimerization; low temperature is preferred. If the monomer has to be stored for more than a few hours, it must be protected against oxygen to prevent peroxidation and polymer formation. Cyclopentadiene monomer reacts spontaneously with oxygen of the air to form brown, gummy peroxide-containing products.

Health and Safety Factors

Toxicological studies conducted on DCPD indicate that it is a moderately toxic material and, to some extent, an irritant and a narcotic. By oral administration in the rat, the LD₅₀ is 0.82 g kg of body weight, and by skin absorption in the rabbit, the LD₅₀ is 6.72 mL kg. An atmospheric concentration of 2000 ppm causes death in rats exposed for a period of 4 hours.

Studies with rats indicate that DCPD is not a myelotoxic chemical; ie, it has no deleterious effects on the blood and blood-forming organs. Toxicologically, it appears to be similar in action to the terpenes rather than to benzene.

The 1978 TLV for DCPD was established at 5 ppm by the ACGIH. This is for a daily 8-hour exposure on a time-weighted average.

Uses of Cyclopentadiene and Dicyclopentadiene

Because of the unusual reactivity of the DCPD molecule, there are a number of wide and varying end use areas. The primary uses in the U.S. are: DCPD-based unsaturated polyester resins (36%); hydrocarbon type resins, based on DCPD alone or with other reactive olefins (39%); EPDM elastomers via a third monomer, ethylenenorbornene or DCPD (16%); and miscellaneous uses (9%), including polychlorinated pesticides, polyhalogenated flame retardants, and polydicyclopentadiene for reaction injection molding (39).

Polyesters. Unsaturated polyester resins based on DCPD, maleic anhydride, and glycols have been manufactured for many years. At least four ways of incorporating DCPD into these resins have been described (45). The resins are mixed with a cross-linking compound, usually styrene, and final polymerization is accomplished via a free-radical initiator such as methyl ethyl ketone peroxide.

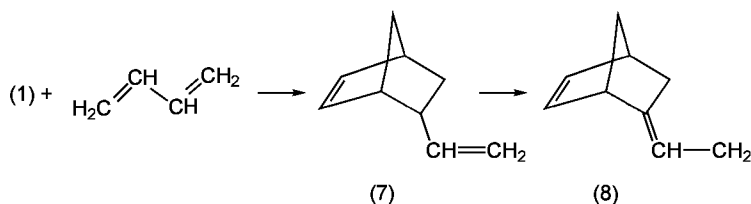
The resins have wide application. In nonreinforced form they serve as insulating coatings for electrical coils (46–47). As fiber reinforced resins, they can be made by reaction injection molding into laminates, castings, and coatings (48–49). Fiber-reinforced resins are used in marine applications (recreational boats); automotive parts (50); bathroom countertops and shower stalls and tubs; and more recently as ballistic protection for military vehicles and aircraft.

Derivatives of CPD have also been incorporated into these resins. CPD and 2-butene-1,4-diol have been condensed in ethanol and catalytically hydrogenated *in situ* to give 2,3-bis(hydroxymethyl)bicyclo[2.2.1]heptane (51). This latter compound is used as a chain extender in polyesters for engineering plastics (52).

Hydrocarbon Resins. Dicyclopentadiene is widely used in both crude and purified form as a monomer in hydrocarbon resin production (see HYDROCARBON RESINS). These resins, produced in both the United States and Europe, are polymerized from concentrated DCPD streams or from the previously mentioned resin oils. The DCPD-containing stream may be polymerized with Friedel-Crafts (qv) (53) catalysts either alone or in admixture with the resin oil (54–55), or with aliphatic olefins and diolefins (56) or by thermal polymerization (57–59).

Producers of hydrocarbon resins buy either high or low purity cyclic streams to make a specific type of resin characterized by softening point, color, unsaturation, etc. In many cases these variables can be controlled or modified as a function of the weight percentage of DCPD present. These products are generally marketed as solids in the form of flakes or pellets in bags, drums, and bulk quantities. The principal end use areas are adhesives, rubber tackification (tires and floor tiles), surface coatings, and inks. Lighter colored resins are generally obtained by cationic polymerization with Lewis acids, especially BF₃. Water white resins are made via hydrogenation of the initially produced polymer and are widely used as tackifiers and sealants.

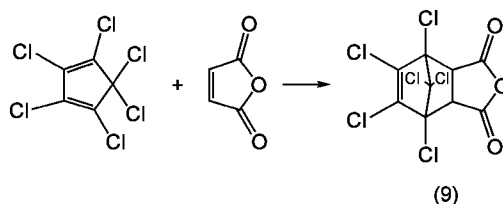
Elastomers. Ethylene-propylene terpolymer (diene monomer) elastomers (EPDM) use a variety of third monomers during polymerization (see ELASTOMERS, ETHYLENE-PROPYLENE DIENE RUBBER). Ethylenenorbornene (ENB) is the most important of these monomers and requires dicyclopentadiene as a precursor. ENB is synthesized in a two step preparation, i.e., a Diels-Alder reaction of CPD (via cracking of DCPD) with butadiene to yield 5-vinylbicyclo[2.2.1]-hept-2-ene [3048-64-4] (7) where the external double bond is then isomerized catalytically toward the ring yielding 5-ethylidenebicyclo[2.2.1]-hept-2-ene [16219-75-3] (ENB) (8) (60).



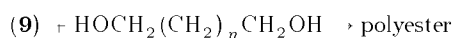
In the EPDM polymerization, the double bond of the bicycloheptene ring system of ENB is involved. The amount of third monomers used in any polymerization varies, but it is usually present at less than 10 wt% of the finished polymer.

Polychlorinated Pesticides. A once substantial but now diminished use for DCPD is in the preparation of chlorinated derivatives for further use or synthesis into pesticide compounds (see INSECT CONTROL TECHNOLOGY). Soil permanence and solubility of the products in human fatty tissues have considerably restricted the use of these compounds. The more prominent chlorinated pesticides were aldrin, dieldrin, chlordane, and heptachlor, all of which use hexachlorocyclopentadiene as a starting material. Aldrin and dieldrin are no longer used in the U.S. Chlordane and heptachlor are still produced, but only for export use.

Flame Retardants. Although the use of chlorinated derivatives of DCPD has been restricted in the pesticide area, some are widely used in flame and fire retardant chemicals (see FLAME RETARDANTS). The starting material is the fully chlorinated DCPD cracked to monomeric hexachlorocyclopentadiene, which is then converted via a Diels-Alder reaction with maleic anhydride to a reactive bicyclic anhydride (9), known as chlorendic anhydride [115-27-5].



This anhydride can then be esterified with polyols yielding a polyester with a residual double bond available for further reaction. Applications for chlorendic anhydride derivatives are in polymers for building materials, paints, and other coatings.



Poly(dicyclopentadiene). The development of polydicyclopentadiene [25038-78-2] for reaction injection molding is an area which has generated much interest. The polyDCPD is obtained via metathesis polymerization of high purity (usually greater than 98%) DCPD. Excellent reviews (61–62) of the chemistry and properties of polyDCPD have been published. The patent literature of polyDCPD synthesis, catalysts, modifiers, and applications is dominated by Hercules (44 patents) and B. F. Goodrich (43 patents) in the U.S. Other participants are Orkem, Shell, Nippon Zeon, and Teijin.

The catalysts are primarily DCPD-soluble derivatives of tungsten and molybdenum and the activators are aluminum alkyls (63–64). Polymerization is accomplished by mixing equal amounts of liquid DCPD (at >32°C), one part of which contains the catalyst and the other of which contains the activator. The mixture is rapidly injected into a mold, where the polymerization takes place. Polymerization times are from under 30 seconds to several minutes, depending on the size of the part, mold temperature, and modifiers added to the polymerizate.

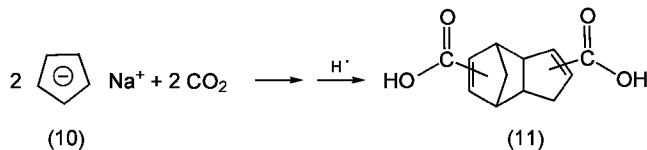
The polyDCPD has good flexural modulus and excellent impact resistance (61). Current uses for polyDCPD are in golf carts, snowmobiles, and automotive bumpers (62). The polymer is viewed as having a high potential, especially in automotive body panel applications.

Specialty Derivatives and Uses

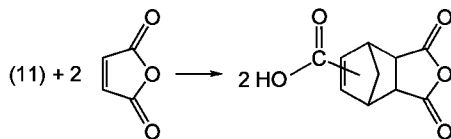
Fuels. Because of their high density, both DCPD and CPD produce high heat when burned. This property would make them excellent high energy fuels except for their reactive characteristics and, therefore, poor storage stability. The tetrahydrogenated products of these reactive diolefins have been claimed to be high energy fuels for racing cars, missiles, etc (65–66) (see FUELS). A current use for the tetrahydrogenated DCPD compounds is as fuel (JP-10) for cruise missiles.

Dicyclopentadienedicarboxylic Acid. Sodium cyclopentadienide (10) can be carboxylated to give dicyclopentadienedicarboxylic acid (11)

in 85–95% yields. The dicyclopentadienedicarboxylic acid can be used in making alkyd resins (qv) that have excellent air-drying and baked-film properties. It can also be converted into many other products characteristic of dibasic acids.

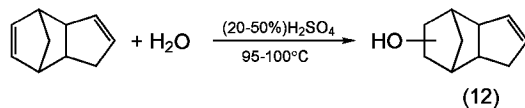


Dicyclopentadienedicarboxylic acid can undergo a Diels-Alder reaction in the presence of a stoichiometric amount of a dienophile at 140–190°C to give an adduct of the monomeric acid. The yield of adduct is usually 75–95%. A large number of polyfunctional compounds are easily prepared in this manner. The reaction with maleic anhydride gives a tribasic acid.



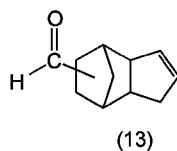
Other Derivatives and Applications. Copolymerization of DCPD with other unsaturated substances has received wide attention, and several useful applications have been developed. With drying oils (qv) thermal copolymerization leads to the production of resinous products, the so-called bodied oils, that give improved drying and result in paint and varnish coatings of greater resistance to weathering.

Numerous reactions of DCPD (and higher oligomers) have been described in the patent literature. These involve addition of the reacting chemical at the double bond of the bicycloheptene ring portion of the DCPD molecule to give such products as secondary alcohols, ethers, esters, and halides. When endo-DCPD is used as the starting material, the resulting products have the exo configuration. For example, in the presence of dilute sulfuric acid, water adds to DCPD to form a secondary alcohol (12). Dicyclopentadiene alcohol can be used as a component of unsaturated polyester resins, perfumes, and plasticizers.



In an analogous manner, DCPD reacts with alcohols and phenols to form ether derivatives, and with halogen acids, thiocyanic acid, and various carboxylic acids to form esters. These esters are used as perfume components (67). Dicyclopentadiene alcohol and a number of the ethers, esters, and glycol adducts have been claimed as coal and ore flotation aids (68).

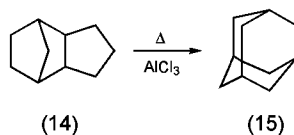
Selective hydroformylation of DCPD has afforded monoformyl derivatives (13), useful as perfume constituents and synthetic rubber intermediates (69) and dimethanolic derivatives useful for the manufacture of polycarbonates and polyesters (70).



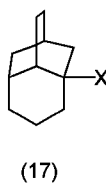
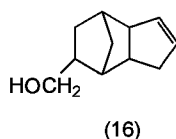
Cyclopentadiene oligomers have been formed by vapor deposition of CPD on kaolin to afford a sorbent for removal of oil from water (71). They are also employed as coatings for controlling release rates of fertilizers (72). Thermal addition of sulfur to a mixture of DCPD and CPD oligomers has led to a number of beneficial applications such as waste water oil adsorbent powdery foams (73), plasticized backing for carpets and artificial turfs (74), and in modified sulfur cements for encapsulating low-level radioactive wastes (75).

Cyclopentadiene itself has been used as a feedstock for carbon fiber manufacture (76). Cyclopentadiene is also a component of supported metallocene—alumoxane polymerization catalysts in the preparation of syndiotactic polyolefins (77), as a nickel or iron complex in the production of methanol and ethanol from synthesis gas (78), and as Group VIII metal complexes for the production of acetaldehyde from methanol and synthesis gas (79).

The synthesis of adamantane (15), tricyclo[3.3.1.1^{3,7}]decane [281-23-2], by heating tetrahydrodicyclopentadiene (14) [6004-38-2] in the presence of aluminum trichloride illustrates another aspect of the synthetic utility of DCPD (80). Adamantane is the base for drugs that control German measles and influenza (80–81) (see ANTIVIRAL AGENTS).



CPD is the starting material for 8-hydroxymethyl-tricyclo[5.2.1.0^{2,7}]decane [57286-50-8] (16) (or octahydro-4,7-methano-1*H*-indene-5-yl-methanol), which is in turn converted into 3-halo-4-homoisotwistane (17), an intermediate for antiviral drugs (82).



The mechanism of the rearrangement is discussed in Reference 83.

BIBLIOGRAPHY

"Cyclopentadiene and Dicyclopentadiene" in *ECT* 1st ed., Suppl. Vol. 2, pp. 282–296, by H. K. Wiese, Esso Research and Engineering Company; in *ECT* 3rd. ed., Vol. 6, pp. 688–704, by H. K. Wiese, Esso Research and Engineering Company; in *ECT* 3rd ed., Vol 7, pp. 417–429, by M. Fefer and A. B. Small, Exxon Chemical Company.

1. A. F. Pame and J. M. Terent'eva, *Usp. Khim* **20**, 560 (1951).
2. J. H. Wells and P. J. Wilson, *Chem. Rev.* **34**, 1 (1944).
3. M. Moulin, *Bull. Assoc. Fr. Tech. Pet.* **135**, 563 (1959).
4. A. S. Onishchenko, *Diene Synthesis*, Old Bourne Press, London, 1964, pp. 274–320.
5. Y. Ishii, *Sekiyu Gakkaishi* **32**, 229–236 (1989).
6. Y. Ishii, *Petrotech (Tokyo)* **11**, 785–787 (1988).
7. G. Xu, Y. Ye, and S. Xu, *Shiyou Huagong* **14**, 175–185 (1985).
8. V. N. Vostrikova and A. A. Grigor'ev, *Khim. Prom-st (Moscow)*, (3), 134–138 (1985).
9. V. N. Vostrikova and T. S. Volkova, *Neftepererab. Neftekhim., (Moscow)*, (7), 26–28 (1983).
10. C. Lou, *Huaxue Tongbao*, (7), 415–421 (1982).
11. H. Komai and A. Ishikawa, *Petrotech (Tokyo)* **3**, 422–427.
12. W. Meyer, *Hydrocarbon Process.* (Sept. 1976).
13. A. H. Wright, *Organomet. Chem.* **16**,351–384 (1987).
14. W. E. Watts, *Organomet. Chem.* **14**, 342–372 (1986).
15. P. M. Maitlis, *Chem. Soc. Rev.* **10**, 1–48 (1981).
16. K. G. Caulton, *Coord. Chem. Rev.* **38**, 1–43 (1981).
17. B. Kanellakoupulos, *NATO Adv. Study Inst. Ser., Ser. C.* **44** (1978) (*Organomet. f-Elem.*), 1–35 (1979).
18. K. C. Ramey and D. C. Lini, *J. Magn. Reson.* **3**, 94 (1970).
19. **U.S. Pat. 4,138,419** (Feb. 6, 1979), M. Arakawa, R. Ohno, K. Ishikawa, N. Yamahara, and H. Matsui (to Japan Synthetic Rubber Co.).
20. P. V. French, L. Roubinek, and A. Wassermann, *Proc. Chem. Soc.* **248**, (1960).
21. *Chem. Eng. News* **38**, 46 (Dec. 19, 1960).
22. J. G. Martin and R. K. Hill, *Chem. Rev.* **61**, 537 (1961).
23. D. R. Eckroth, *J. Org. Chem.* **41**, 394 (1976).
24. K. Alder and G. Stein, *Ann.* **504**, 205 (1933).
25. J. Pirsch, *Ber.* **67B**, 1115 (1934).
26. K. Alder and G. Stein, *Ber.* **67B**, 613 (1934).
27. K. Alder and co-workers, *Ann.* **496**, 204 (1932).
28. G. Becker and W. A. Roth, *Ber.* **67B**, 627 (1934).
29. E. T. McBee and H. E. Ungnade, *Chem. Rev.* **58**, 249 (1958).
30. N. A. Milas and L. S. Maloney, *J. Am. Chem. Soc.* **62**, 1841 (1940).
31. R. Criegee, *Ann. Chem.* **481**, 263 (1940).
32. P. Seguin, *Compt. Rend.* **216**, 667 (1943).
33. G. O. Schenk and D. E. Dunlap, *Angew. Chem.* **68**, 248 (1956).
34. **U.S. Pat. 3,644,430** (Feb. 22, 1972), K. H. Schulte-Elte (to Firmenich et Cie).
35. C. G. Venier and E. W. Casserly, *Am. Chem. Soc. Div. Petr. Chem.* **35**, 260–264 (1990).
36. E. W. Casserly and C. G. Venier, *Am. Chem. Soc. Div. Petr. Chem.* **35**, 265–271 (1990).
37. S. M. Csicsery, *J. Org. Chem.* **25**, 519 (1960).
38. **U.S. Pat. 2,831,904** (Apr. 22, 1958), R. W. R. Krept (to Shell Development Co.).
39. *Chem. Mark. Rep.* (May 14, 1990).
40. A. R. Bujold, *Chemical Economics Handbook*, section 640.5000, SRI International, Menlo Park, Calif., 1989, and private communication with author, 1992.
41. J. Bermejo, C. G. Blanco, M. A. Diez, and M. D. Guillen, *Chromatographia* **23**(1), 33–37 (1987).
42. **U.S. Pat. 3,719,718** (Mar. 6, 1973), F. Grude, K. Halcour, W. Schwerdtel, W. Swodenek, and P. Woernle (to Farbenfab Bayer A.G.).
43. **Jpn. Pat. 61,227,537** (Oct. 9, 1986), M. Usui, K. Kimura, S. Yamamoto, and K. Takashi (to Sumitomo Chemical Industries K.K.).
44. **Jpn. Pat. 60,116,634** (June 24, 1985), E. Kudo and S. Minegishi (to Hitachi Chemical K.K.).
45. P. L. Smith, C. W. McGary, Jr., and L. R. Comstock, *Proc. Annual Tech. Conf. SPI, Reinf. Plast. Div. 22d Annual Meeting*, Washington, D.C., Jan. 31–Feb. 3, 1967, pp. 1–12.
46. **U.S. Pat. 4,623,696** (Nov. 18, 1986), D. W. Mabrey and J. M. Dye (to P. D. George Co.).
47. **U.S. Pat. 4,535,146** (Aug. 13, 1985), A. Mekjian (to U.S. Steel Corp.).
48. **U.S. Pat. 4,532,297** (July 30, 1985), H. C. Gardner (to Union Carbide Corp.).
49. **U.S. Pat. 4,522,977** (June 11, 1985), H. C. Gardner (to Union Carbide Corp.).
50. **U.S. Pat. 4,673,706** (June 16, 1987), K. E. Atkins (to Union Carbide Corp.).
51. **U.S. Pat. 4,948,866** (Aug. 14, 1990), K. C. Liu and S. B. Nelson (to GAF Chemical Corp.).
52. **U.S. Pat. 4,927,976** (May 22, 1990), K. C. Liu (to GAF Chemical Corp.).
53. **U.S. Pat. 3,023,200** (Feb. 27, 1962), M. E. Epstein and N. C. Gangemi (to Pennsylvania Industrial Chemical Corp.).
54. **U.S. Pat. 4,558,107** (Dec. 10, 1985), M. L. Evans and A. B. Small (to Exxon Research & Engineering Co.).
55. **Jpn. Pat. 61,197,616** (Sept. 1, 1986) (to Toho Chemical Industries Ltd.).
56. **Ger. Offen. 2,014,424** (Dec. 23, 1970), H. L. Bullard, R. A. Osborn, and D. R. St. Lyr (to Goodyear Tire & Rubber Co.).
57. **U.S. Pat. 3,084,147** (Apr. 2, 1963), L. P. Wilks (to Velsicol Chem. Corp.).
58. **Eur. Pat. 149,909** (July 31, 1985), B. Bossaert, J. Lechat, A. Lepart, V. L. Hughes, and R. C. C. Yap (to Exxon Research & Engineering Co.).
59. **Jpn. Pat. 63,260,913** (Oct. 27, 1988), S. Hayashida, K. Kimura, Y. Shimada, and E. Takahashi (to Maruzen Petrochemical K.K.).
60. **U.S. Pat. 3,347,944** (Oct. 17, 1967), H. E. Fritz, K. E. Atkins, and G. L. O'Connor (to Union Carbide Corp.).
61. R. P. Geer, *Plast. Eng.* **41**(11) 41–49 (1985).
62. D. S. Breslow, *ChemTech* **20**, 540–544 (Sept., 1990).
63. **U.S. Pat. 4,426,502** (Jan. 17, 1984), R. J. Minchak (to B. F. Goodrich Co.).
64. **U.S. Pat. 4,400,340** (Aug. 3, 1983), D. W. Klosiewicz (to Hercules Inc.)
65. **U.S. Pat. 3,381,046** (Apr. 30, 1968), C. A. Cohen and C. W. Muessig (to Esso Research & Engineering Co.).
66. **U.S. Pat. 3,002,829** (Oct. 3, 1961), J. J. Kolfenbach and H. K. Wiese (to Esso Research & Engineering Co.).
67. **Eur. Pat. 249,005** (Dec. 16, 1987), W. Kleinhama (to Huls A. G.).

68.

U.S. Pat. 4,925,599 (May 15, 1990), M. Biermann, R. Koster, and H. Eierdanz (to Henkel Kommandtgesellschaft auf Aktien).

69.

U.S. Pat. 4,289,913 (Sept. 15, 1981), H. Bahrmann, B. Cornils, and J. Weber (to Bayer A.G.).

70.

U.S. Pat. 4,262,147 (Apr. 14, 1981), P. E. Garrou and G. E. Hartwell (to the Dow Chemical Co.).

71.

USSR Pat. 1,344,401 (Oct. 15, 1987), V. A. Diner, V. N. Cheshchevo, and V. A. Polushkin.

72.

U.S. Pat. 4,678,710 (July 7, 1987), S. Sakimoto, K. Fujita, Y. Yamauchi, and H. Matsue (to Showa Denko K.K.).

73.

Jpn. Pat. 49,060,323 (June 12, 1974), M. Chono, K. Toyomoto, S. Nogami, and N. Matsuo (to Asahi Chemical Industries K.K.).

74.

U.S. Pat. 4,012,545 (Mar. 15, 1973), W. G. Toland (to Chevron Research Co.).

75.

P. D. Kalb, J. H. Heiser, III, and P. Columbo, *Summary and Proceedings of The Twelfth Annual U.S. DOE Low-Level Waste Management Conference*, Chicago, Ill., Aug. 28–29, 1990, pp. 113–125.

76.

Jpn. Pat. 61,282,426 (Dec. 12, 1986), K. Nakamura (to Asahi Chemical Industries K.K.).

77.

U.S. Pat. 4,892,851 (Jan. 9, 1990), J. A. Ewen and A. Razavi (to Fina Technology Inc.).

78.

U.S. Pat. 4,434,248 (May 4, 1982), M. T. Barlow (to British Petroleum Ltd.).

79.

P. von R. Schleyer, *J. Am. Chem. Soc.* **79**, 3292 (1957).

80.

Chem. Weeks **94**, 119 (Apr. 25, 1964); A. W. Galbraith and co-workers, *Lancet ii* **113** (1971).

81.

R. C. Fort, Jr. and P. von R. Schleyer, *Chem. Rev.* **64**, 277 (1964).

82.

Jpn. Pat. 52,073,846 (June 21, 1977), Y. Inamoto, Y. Fujikura, and H. Ikeda (to Kao Soap K.K.).

83.

Y. Inamoto, N. Takaishi, K. Tsuchihashi, K. Aigami, and Y. Fujukura, *J. Org. Chem.* **41**, 771 (1976).

Michael J. Keenan
Exxon Chemical Company

CYTOKININS.

See GROWTH REGULATORS, PLANT.

DAIRY PRODUCTS.

See MILK AND MILK PRODUCTS.

DAIRY SUBSTITUTES

Dairy products (see MILK AND MILK PRODUCTS) have been staple items of the diet for many centuries, and have long been the target for imitation. The development of nutritional guidelines emphasizing the need to reduce total dietary fat, dietary cholesterol [57-88-5], C₂₇H₄ O, and saturated fatty acids (see FATS AND FATTY OILS; FAT SUBSTITUTES), has increased the interest in imitation dairy foods. However, with the exception of butter and cream the market penetration of dairy substitutes has been limited.

The origin of substitute dairy foods extends back to the 19th century with respect to the manufacture of oleomargarine and filled cheese. The number and complexity of these substitutes has increased with increasing technology and improved knowledge of the functionality of the various food ingredients. Generally products made to imitate dairy foods require the same technology, processes, and equipment required for dairy products. The production and marketing of dairy substitutes reflect technology, economics, legislation, politics, nutrition, and public food habits. The relative significance of these factors varies from country to country and within regions of the same country. Market penetration of imitation dairy products is probably greatest in the United States.

Advances in processing technology occurring since the early 1940s, eg, the widespread use of homogenization, fluid blending, and continuous processes, have markedly advanced the development of dairy substitutes. It is relatively simple to manufacture physically stable products from a wide range of fats, proteins (qv), and carbohydrates (qv). Moreover, growth in use of convenience foods and decreased in-home preparation of meals have made the U.S. consumer more willing to change food habits and patterns. In almost all cases, synthetic dairy products are offered at lower cost. Thus economics is perhaps the most important single factor in the initial acceptance of imitation dairy foods. Because of improved technical properties, ability to meet consumer requirements, and an almost unlimited shelf life, many dairy substitutes are preferred over the alternative milk products (1).

Some countries, such as Germany, have laws that restrict imitation dairy products. Products that simulate milk and other dairy food of recognized nutritional value generally are required to be nutritionally equivalent to the dairy products that they imitate. In the United States, where legal standards exist for many substitute dairy products, the laws are less restrictive.

A code of principles accepted by 71 countries has been developed for consumer protection and fair practice in the trade of milk and milk products. Mainly the precise usage of the term milk and terms for different milk products is ensured. Confusion arising from the substitution of milk and milk products with nonmilk fats and or nonmilk proteins is thus avoided. The use of misleading names and information for products that are not milk products is prohibited. Essentially, any product that resembles a dairy product is an imitation or substitute (synthetic) product.

There are no universally accepted definitions of substitute dairy foods, which are referred to as imitations, simulates, substitutes, analogues, and mimics and are associated with terms such as filled, nondairy, vegetable nondairy, and artificial milk, cheese, etc. The term nondairy has been used indiscriminately to describe both imitation dairy products and products legally defined as not being imitation dairy products. Dairy substitutes can be divided into three types: those in which an animal or vegetable fat has been substituted for milk fat; those that contain a milk component, eg, casein [9000-71-9] or whey protein; and those that contain no milk components (see MILK AND MILK PRODUCTS). The first two types make up most of the substitute dairy products.

The International Dairy Federation (IDF) has defined the following terms (1,2). (1) Substitute dairy foods. A substitute dairy product is a food stuff where the intended use is to substitute for milk or dairy foods that have legal standards of identity. In some countries the term substitute requires that the food be nutritionally equivalent to the product for which it substitutes. (2) Imitation dairy foods. An imitation dairy food is a substitute dairy product in which the general composition, appearance, characteristics, and intended use is similar to milk or to a dairy food that has a legal standard of identity. Filled milk products are included under this definition. (3) Synthetic products. These are made in semblance of a dairy product and contain no milk component, although some include sodium caseinate [9005-46-3]. Sodium caseinate is classified as non-dairy because of its industrial usage. (4) Filled. Filled imitation dairy products are made in semblance of a dairy product. A vegetable or animal fat is used to substitute for milk fat. (5) Nondairy. Milk–protein-based imitation dairy products are made to resemble a dairy product and contain one or more proteins derived from milk. Nondairy in this case has previously referred to imitation dairy products that contain proteins derived from casein because the FDA has ruled that such proteins are chemicals and cannot be considered dairy products. More recently, the term non-dairy has been used for products that contain whey proteins derived from milk, using the same rationale as for casein. There has not been a clear ruling from the FDA on this issue as of this writing.

Ingredients

The characteristics and stability of imitation dairy products depend largely on the characteristics of the main ingredients, ie, fat, protein, and carbohydrates, together with the minor functional ingredients that stabilize the fat and protein systems. The ingredients generally include fats or oils and their stabilizing emulsifiers, proteins and their stabilizing gums (qv), and salts and carbohydrates (see VEGETABLE OILS). Many substitute dairy products have been reformulated in the past few years to meet consumers' interest in lower fat foods.

Fats and Oils. In industrial use, fats are considered solid and oils liquid at room temperature. The latter are liquids because of a higher content of unsaturated fatty acids. Hydrogenation converts oils to fats. Fats and oils consist mainly of the triglyceride–rich lipid fraction, which contains small amounts of other lipids, eg, sterols and phospholipids. For imitation dairy products, fats are generally selected to have low and narrow melting point ranges, generally 32–36°C. The fats and oils are characterized in a number of ways, including iodine number, polyunsaturated fatty acid content, saturated fatty acid content. Wiley melting point, and by solid–fat content at different temperatures. The iodine number is an index of the number of unsaturated bonds in the fatty acid molecules of the fat or oil. The polyunsaturated fatty content is given as a percent of the fatty acids having two or more unsaturated bonds. The Wiley melting point is the temperature at which all of the fat or oil is in a fluid state. The solid–fat content is given as the percentage of the fat or oil that is solid at a given temperature as measured by nmr, and is generally measured over the temperature range of 10 to 40°C.

The desired characteristics of a fat or oil used in imitation dairy foods are bland flavor, low peroxide value and good flavor stability, low level of free

fatty acids, resistance to hydrolysis, and a desired solid fat content over the use temperature range of the product. The common fats used in substitute dairy foods are hydrogenated coconut, cottonseed, soybean, groundnut, palm kernel, and various blends of these products. Fat systems having lower melting points are generally preferred as these have a better texture or a better feel to the mouth. Specialized products may contain polyunsaturated oils including corn, rapeseed, safflower, and sunflower oils. Concern about dietary fat has resulted in the use of fats having lower levels of saturated fats as well as a reduction in the amount of total fat in the product.

The physical characteristics of a fat or oil for imitation dairy products are not necessarily dictated by the fat being replaced, but by the composition, processing methods, and conditions of use of the substitute product. Thus, the selection of the fat or oil is generally developed experimentally. The chemical and physical nature of the components of the system, order of addition, shear input, and processing temperature dictate the final interactions and the nature of the product. A comparison of the characteristics of milk fat and three different fats that are used in satisfactory filled and imitation milks are listed in Table 1 and those that comprise whipped topping, in Table 2.

Table 1. Fat Characteristics for Imitation and Filled Milks^a

Characteristics	Milk fat	Fat products		
		A	B	C
SFI, ^b %				
at 10°C	34	25	15	15
at 21°C	13	13	11	9
at 27°C	8	9	10	8
at 33°C	0.5	4	6	2.5
iodine value	33	80	95	105
polyunsaturated fatty acid, wt % (cis-cis)	1.6	21	24	36
saturated fatty acids, wt %	65.5	21	20	18

^a Ref. 2.
^b SFI = solid fat index.

Table 2. Fat Characteristics for Imitation Dairy Products^a

Characteristic	Ice cream and whipped topping				Coffee whiteners	
	A	B	C	D	Fluid	Dry
SFI, ^b %						
at 0°C	38	30	24	25	25	61
at 21°C	23	20	13	18	14	50
at 27°C	20	19	10		9	46
at 33°C	12	14	6	4	4	28
at 41°C	4	8	2			
iodine value	68	82	83	85	80	67

^a Ref. 2.
^b SFI = solid fat index.

Fat Replacers. The reduction of fat in substitute dairy products results in an increase in water and a stress on the food system both in respect to body and texture, and to flavor. There is no universal fat replacer, but microparticulated proteins having particle sizes <10 μm and or starch derivatives, and gums have been used as fat replacers.

Emulsifiers. An emulsion is a two-phase system consisting of two immiscible liquids, the one being dispersed as fine globules in the other (see EMULSIONS). In food systems, two general types of emulsions exist: oil-in-water and water-in-oil. An emulsifier is a product that has a water-soluble and a fat-soluble portion in the same molecule and stabilizes fat containing food products. Emulsifiers may be natural products, eg, phospholipids and proteins, or derived from natural products, eg, esters of long-chain fatty acids and a polyhydric alcohol (see ALCOHOLS, POLYHYDRIC). Types of emulsifiers used in dairy substitutes include: mono- and diglycerides, lactic acid [598-82-3] and fatty acid esters of glycerol [56-81-5], polyglycerol esters of fatty acids, sorbitan esters of fatty acids, polyoxyethylene esters of fatty acids, and propylene esters of fatty acids. Frequently a combination of emulsifiers are used to achieve the desired characteristics in the final product. The emulsifiers should provide for a stable emulsion resistant to coalescence and breaking, be bland and contribute no off-flavors, provide an emulsion that is resistant to creaming or separation, and have an optimum hydrophile-lipophile balance for the given systems. All imitation dairy foods are oil-in-water emulsions except for margarine which is the only product that is a water-in-oil emulsion.

The hydrophile-lipophile balance (HLB) is an empirical system based on the fact that oil-water (o/w) emulsions are best stabilized by water-soluble-emulsifiers and water-oil (w/o) emulsions are best stabilized by oil-soluble ones (3). The HLB scale runs from 0-20 and is based on the ratio of the saponification number of ester, *S*, to the acid number of recovered acid, *A*, where $HLB = 20(1 - S/A)$. The dispersibility of an emulsifier in water is related to HLB value.

Dispersibility	HLB value
no dispersibility in water	1-4
poor dispersibility	3-6
milky dispersion after shaking	6-8
stable milk dispersion	8-10
translucent-to-clear dispersion	10-13
clear solution	13+

The application of surfactant is also related to HLB value.

Application of surfactant	HLB value
w/o emulsifier	4-6

wetting agent	7–9
o/w emulsifier	8–18
detergents	13–15
solubilizers	10–18

Table 3 gives HLB values of some of the important emulsifiers. The HLB optimum for a given emulsifier varies with the components of the food system. A coconut oil–water emulsion that shows optimum stability with an HLB of 7–9 shows a shift in requirements for stability upon addition of casein and electrolytes to an optimum stability using an emulsifier having an HLB of 3–5. In addition, the stability of an emulsion can be affected by the chemical nature of the emulsifier. The optimum HLB for an emulsifier in a given system is influenced by the other ingredients as is illustrated for a model synthetic milk system in Figures 1 and 2.

Table 3. Emulsifiers and their Characteristics

Emulsifier	Melting point, °C	Viscosity, ^a mPas(cP)	HLB rating
mono- and diglycerides			
(48–59° o monoglycerides)	46–53	ca 150	2.8
(61–69° o monoglycerides)	52–61	ca 150	3.2–3.5
sorbitan monostearate	53		4.7
sodium stearoyl-2-lactylate			4.5
polyoxyethylene (20) monostearate		600	14.9
polyoxyethylene (20) tristearate	ca 33	500	10.5
polyoxyethylene (20) monooleate ^b		400	15.0
mono- and diglycerides (40° o) + polysorbate 80 (60° o)	54–47		5.9

^a At 25°C
^b Also known as polysorbate 80.

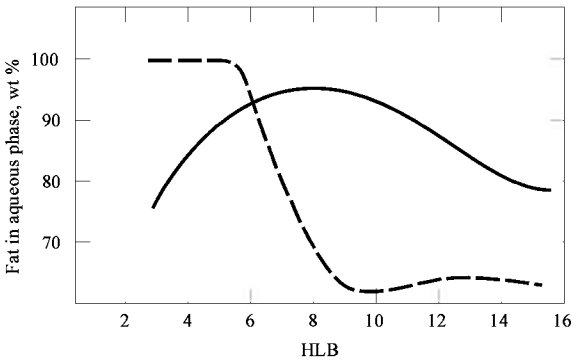


Fig. 1. Effect of system components on emulsifier HLB. (—), fat–caseinate–salts–water–emulsifier; (---), fat–water–emulsifier (4).

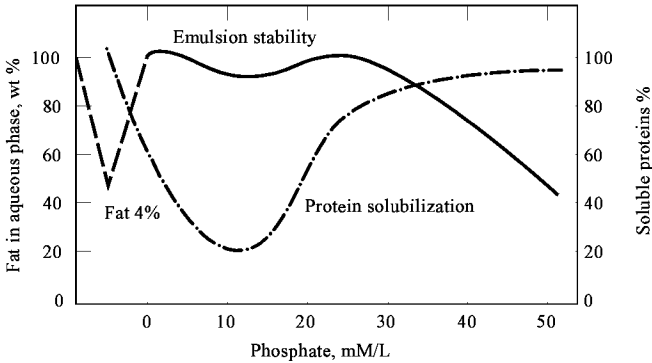


Fig. 2. Effect of changes in system components on (—), emulsion stability and (---), protein solubilization in an imitation milk where the protein and fat are present at 2 and 4 wt % o, respectively (4). The concentration of calcium is constant at 25 mM L for all values of phoshate; () represents the system prior to the addition of calcium.

Generally, emulsifiers are selected to minimize coalescence of globules. However, in the case of high fat whipped products, eg, ice cream and whipped toppings, the emulsifier provides for controlled agglomeration of fat globules, thus giving the desired body to the final product. The agglomeration helps to control air micelle size and stability.

The physical form of the food relates to the selection of the emulsifier. For example, coffee whiteners are marketed as liquids, frozen products, or dry powders. Optimum emulsifier differs for each system. For liquid products, the emulsifier must impart good stability to the emulsion so that it remains in a uniform state on standing after preparation and prior to sale. It must also prevent oiling off (emulsion breaking) and feathering (formation of unstable white flocs on top that resemble feathers) when used in coffee. For the frozen product, the emulsifier must also contribute to the freeze–thaw stability of the emulsion. In dry products, the emulsifier must contribute to the free-flowing properties and to the dispersal of the product in coffee. The dry whitener should sink to the bottom of the coffee and disperse rapidly. The selection of emulsifier differs for each product. The lowest HLB emulsifiers are used for liquid products and the highest HLB emulsifiers are used for dry products. Thus, mono- and diglycerides are commonly used in liquid coffee whiteners and blends of the emulsifiers are used for dry products. A typical blend contains mono- and diglycerides, polyoxyethylene esters, and sorbitan fatty acid esters. The type and level of emulsifier also depends on the type and concentration of protein used in the product because the protein also contributes to the

emulsification of the fat. A common concentration of emulsifier in imitation dairy foods is 5% of the fat concentration.

Proteins. Proteins are especially significant in imitation dairy foods with respect to nutritional and physical properties of the product (5). The relative significance of the nutritional quality of the protein depends upon product type and the extent to which the product contributes to the total protein intake of a given population. Thus, protein quality is extremely important in imitation milk, significant in imitation cheese, and less significant in coffee whiteners and whipped toppings.

The role of protein with respect to physical properties varies with the type of product. Proteins contribute to a number of functions in an imitation food. These include emulsification, gelation, melting, water binding, and whipping.

Factors to be considered in protein selection include: nutritional quality required, specific functionality required, solubility dispersibility (ease of incorporation into formulation), good flavor quality (no off-flavor contributed to product), and stability to processing conditions (see also [FOOD PROCESSING](#)). In addition to emulsification, which is important to all products, the desired functionality of proteins in substitute cheese is melting; in coffee cream, whitening and buffering; in cultured products, gelation; in ice cream, whipping; in milk-based pudding, gelation; and in whipped topping, whipping.

A wide number of protein sources are available for use in dairy substitutes. These include: animal proteins, ie, skim milk in liquid, condensed, or dry form (filled products); casein, caseinates, and coprecipitates; whey proteins; oil-seed proteins, fish proteins; and blood proteins. Oil-seed protein sources include soybean protein concentrates and isolates, groundnut protein, cottonseed protein, and sunflower seed, rapeseed, coconut, and sesame seed proteins (see [SOYBEANS AND OTHER OIL SEED](#)). Other sources are leaf and single-cell proteins (see [FOODS, NONCONVENTIONAL](#)). Of these protein sources, milk and soybean proteins are most widely used. Protein usage is based on economics, flavor, functionality, and availability.

Casein and Caseinates. Definitions for edible casein and edible caseinate have been suggested (5). Edible casein is the product obtained by separating, pressing, and drying lactic- or mineral-acid-precipitated coagulum of skimmed milk. Edible caseinate is the product obtained by drying aqueous solutions prepared by combining dry edible casein or fresh edible casein curd and food-grade alkali. Casein is also made from rennet coagulation of skim milk. Casein and caseinates are the most widely used proteins in substitute dairy products, primarily because of superior functionality and flavor as compared to most vegetable proteins. Generally, the salt form of casein is used and the two common forms are sodium caseinate and calcium caseinate [9005-43-0]. Sodium, potassium, ammonium, and calcium caseinates are available.

Proposed IDF standards for caseinate are listed in Table 4. In most cases the sodium salt is preferred for emulsification; the calcium salt is preferred for imitation cheese. Casein and caseinates must be stored carefully and evaluated for flavor before use in products. Improperly manufactured or stored casein-caseinate has a very strong, musty off-flavor. Excessive fat content, high lactose and moisture contents, and high storage temperatures contribute to undesirable flavor development.

Table 4. IDF Caseinate Specifications^a

Property	IDF standard 72, edible caseinate	
	Extra grade ^b	First grade ^b
color	white-to-pale cream	white-to-pale cream
flavor and odor	bland	bland
moisture, wt %	6.0	8.0
protein (dry basis N × 6.38), ^c wt %	90.0 ^d	88.0 ^d
fat (dry basis), wt %	1.5	1.5
lactose, wt %	0.5	1.0
scorched particles, mg/g	15.0–25	22.5–25
extraneous matter, 50 g ⁻¹	^e	^e
copper, ppm	5	5
lead, ppm	5	5
iron, ppm	20	20
total plate count, g ⁻¹	10,000	100,000
coliforms, 0.1 g ⁻¹	^f	^f
mold and yeasts, g ⁻¹	50	50
thermophilic organisms, g ⁻¹	5,000	5,000

^a Ref. 6.
^b Values given are maximum values unless otherwise indicated.
^c N is total nitrogen.
^d Values given are minimum values.
^e Free from extraneous matter.
^f Negligible value.

Whey Proteins.

Whey is the fluid obtained by separating the coagulum from cream and or skim milk, and is a by-product of either casein or cheese manufacture. The composition of whey is determined by the method of curd formation, curd handling practices, and methods of handling whey as it is separated from the curd. Dried acid whey contains ca 12.5 wt % protein (total nitrogen × 6.38), 11.0 wt % ash, and 59 wt % lactose, whereas sweet whey contains 13.5 wt % protein, 1.2 wt % fat, 8.4 wt % ash and 74 wt % lactose. The composition varies with the type of acid used (7).

Whey has been used in some substitute dairy products but not as a source of protein. Whey proteins have been used in dairy substitutes only since the commercialization of ultrafiltration (qv) technology. Membranes are used that retain protein and permit water, lactose, and some minerals to pass through as permeate. Protein concentrates are available from both acid and sweet whey and in concentrations of 35–80 wt % protein. Whey protein isolates are commercially available having protein >90 wt%. The cost of these isolates is too high, however, to make them economical for substitute dairy foods.

Functionality of whey protein concentrates varies with whey type and concentration. Table 5 gives compositional data for whey protein concentrates from different sources of whey. These concentrates are used in a limited number of products; ice cream and other frozen desserts, fermented products, coffee whiteners, and whipped toppings.

Table 5. Whey Protein Concentrate Compositions from Different Sources^{a, b}

Property	Rennet	Lactic acid	Sulfuric acid	Cheese
TN × 6.38 ^c	79.06	80.00	80.00	84.84
nonprotein nitrogen (NPN)	0.78	0.35	0.10	0.63
true protein	73.96	77.77	79.36	80.82

lactose	10.70	10.00	9.30	2.33
ash	3.30	2.5	2.5	2.63
fat	3.30	5.0	5.0	7.96
PO ³ ₄	0.22	0.60	0.42	0.11
Ca ²⁺	0.53	0.42	0.37	0.60
SO ² ₄	0.06		0.75	0.15
Mg ²⁺	0.005	0.010	0.025	0.033
Na ⁺	0.28	0.16	0.16	0.58
K ⁺	0.69	0.21	0.21	0.17
Cl		0.30	0.24	0.52
lactate		1.50		0.45

^a Ref. 7.
^b Values are given as % of total solids.
^c TN is total nitrogen in the protein.

A significant problem is the wide variation in properties in the same type of product both from different manufacturers and from the same source. More recently there has been vast improvement in the consistency of these products from a given manufacturer and usage in dairy substitutes is expected to expand. These isolates do have a higher protein efficiency ratio (PER) than casein and are equal to egg protein.

Whey proteins that have been heat precipitated under very high shear have a particle size between 1 and 3 micrometers, and give the impression of fat in some products. These microparticulated whey proteins are being used as fat replacers in frozen desserts and processed cheese substitutes.

Soybean Proteins. Based on protein content, there are three main groups of soybean protein products: flakes, flours, and grits, each having a minimum protein content of 40 wt %; concentrates having a protein content of 70% on a dry weight basis; and isolates containing 90% or more protein on a dry weight basis. For imitation products, the concentrates and isolates are the more important. The flavor quality of the products improves with increased protein content but only isolates approach a flavor quality sufficient for most imitation dairy foods when used at the same protein level as milk proteins in substitute products. Additional processing steps can further improve flavor, but at considerable expense. Continued improvement in flavor of soy protein products can be expected.

Soy Protein Concentrates. Soy protein concentrates are produced from dehulled soybeans by removing the majority of the oil and water-soluble nonprotein components. These contain not less than 70% protein on a dry weight basis (1). They are obtained by leaching oligosaccharides, minerals, and other soluble constituents by one of several methods: (1) aqueous acids at pH 4–5, (2) hot water, (3) 20–80% aqueous organic solvents, or (4) chilled water. The physical properties vary according to the method used. Concentrates prepared by acid leaching and pasteurization in the absence of heat treatment have a higher soluble protein content.

Soybean Protein Isolates. Soybean protein isolates, having a protein content of >90 wt%, are the only vegetable proteins that are widely used in imitation dairy products (1). Most isolates are derived from isoelectric precipitation, so that the soybean protein isolates have properties that are similar to those of casein. They are insoluble at their isoelectric point, have a relatively high proportion of hydrophobic amino acid residues, and are calcium-sensitive. They differ from casein in that they are heat-denaturable and thus heat-labile. The proteins have relatively good nutritional properties and have been increasingly used as a principal source of protein. A main deterrent to use has been the beany flavor associated with the product. Use is expected to increase in part because of lower cost as compared to caseinates. There has been much research to develop improved soybean protein isolates.

Soy protein isolates are derived from defatted flakes or flours by removing water-insoluble polysaccharides, oligosaccharides, and other low molecular weight components (1). The flakes or flour are extracted with an aqueous alkaline medium (pH 7–8.5). The clarified extract is acidified to pH 4.5 with food-grade acid. The precipitated proteins are separated by centrifugation and then are washed, neutralized, and dried as the proteinate. The pH treatment, temperature of precipitation, and ionic strength dictate the ratio of the protein classes (glycinin and conglycinin) and the functionality of the isolates. Glycinin and conglycinin vary in functionality, and variations in the ratios of these are significant in the behavior of the soy protein isolate in substitute dairy. Some soy protein isolates are slightly hydrolyzed to improve solubility and emulsifying properties. Blends of soy with either caseinates or whey proteins have been prepared to extend soy protein functionality and to minimize off-flavor problems. Compositions of soy protein concentrates and isolates are given in Table 6.

Table 6. Analyses of Commercial Soy Protein Concentrates and Isolates^a

Analysis	Concentrates ^b			Isolates ^b			
	A	B	C	A	B	C	D
moisture, wt %	6.7	5.2	3.1	4.7	6.4	7.6	3.7
protein, wt %	66.2	67.3	69.6	92.8	92.2	92.9	94.7
protein (dry basis), wt %	70.9	71.1	72.2	97.4	98.7	100.0	98.4
fat	0.3	0.3	1.2				
crude fiber, wt %	3.5	3.4	4.4	0.2	0.1	0.1	0.2
ash, wt %	5.6	4.8	3.7	3.8	3.5	2.0	2.7
nitrogen solubility index	5.0	69.0	3.0	85.0	95.0		
pH (1:10 aqueous dispersion)	6.9	6.6	6.9	7.1	6.8	5.2	5.5

^a Ref. 1.
^b A, B, C, D each represent a different product.

Other Proteins. Groundnut, fish, and cottonseed proteins have been used to a limited degree in dairy substitutes. The properties of the materials are discussed in the literature (5).

Gum Stabilizers. Gums (qv) are used in substitute dairy products for one of the following reasons (3): to provide viscosity control and improve mouth feel; to improve whipping (aerating) properties of whipped products; to provide a protective colloid to stabilize proteins to heat processing; to modify the surface chemistry of fat surfaces to minimize creaming; to provide acid stability to protein systems; to increase freeze–thaw stability; to provide a hard-water resistant coffee whitener; and to provide desired melting characteristics to imitation cheese. Gums can be classified as neutral and acidic, straight- and branched-chain, gelling and nongelling. The principal gums and the functions in imitation dairy products are listed in Table 7. A listing of gum applications in dairy foods that apply to imitation products is given in Table 8.

Table 7. Gums and Associated Functions in Dairy Substitutes

Gum	Imitation product use	Function
alginate	ice cream	aeration, reduce whip time
	cheese	texture modifier, prevent oil separation
carrageenan	milk puddings	gelation

locust bean gum	ice cream	prevent wheying off
	infant formula	stabilize proteins to heat
	evaporated milk	stabilize proteins to heat
	ice cream	smooth melt down, freeze–thaw resistance
guar gum	cheese spread	texture modifier
	ice cream	water binding, body control
carboxymethyl	ice cream	prevent ice crystal growth
cellulose	whipped topping	aeration, protective colloid
xanthan gum	sour cream	prevent wheying off

Table 8. Application of Gums in Dairy Products^a

Product	Agar	Algin	Carra-g eenan	Furcel-l aran	Arabic	Xan -tha n	Traga- canth	Hydroxy- propyl cellulose	Locust bean	Guar	CMC ^b	Methyl cellulose	Karaya
ice cream stabilizer	X	X	X	X	X		X		X	X	X	X	X
ice milk	X	X	X	X	X		X		X	X	X	X	X
milkshake	X	X	X				X		X		X		
sherbets	X	X	X	X	X		X		X	X	X		X
chocolate milk drink		X	X	X	X	X	X		X	X	X		
pudding	X	X	X	X	X	X			X	X			X
cream cheese	X	X	X		X				X	X			
cheese spread		X	X		X	X	X	X	X	X	X		X
Neufchâtel process cheese	X	X	X		X		X		X	X			X
whipped cheese		X	X			X		X	X				X
yogurt	X		X		X				X	X			

^a Ref. 3.
^b CMC = carboxymethyl cellulose.

Gums are generally used in concentrations of 0.02–0.5 wt % or 1–6 wt % of the protein level in the food. Because of differing functionalities, combinations of gums provide a better product than an individual gum.

Stabilizing Salts. Citrates and phosphates are used in substitute dairy products for one or more of the following purposes (3): to alter buffering capacity of the system; to improve the stability of the protein to calcium ions; to improve the heat stability of the protein; to minimize the age gelation of ultra-high temperature-(UHT) processed substitute products; to serve as emulsifying salts in imitation cheese manufacture; and to modify the water-binding capacity of proteins and improve solubility. Common phosphate salts used in these foods are monosodium phosphate [7558-80-7] (MSP), disodium phosphate [7558-79-4] (DSP), trisodium phosphate [7601-54-9] (TSP), disodium pyrophosphate [7558-16-9], tetrasodium pyrophosphate [7722-88-5] (TSPP), sodium tripolyphosphate [7758-29-4] (STP), Na₅O₁₀P₃, sodium hexametaphosphate [10124-56-8] (SHMP), sodium trimetaphosphate [7785-84-4], Na₃O₉P₃, and sodium tetrametaphosphate [13396-41-3] (3). Although there are similarities in some functional uses, phosphates are more broadly used than citrates in substitute dairy foods.

Phosphates, which react with calcium to reduce the calcium ion activity, assist in stabilizing calcium-sensitive proteins, eg caseinate and soy proteinate, during processing. Phosphates also react with milk proteins. The extent of the reaction depends upon chain length. Casein precipitates upon addition of pyrophosphates, whereas whey proteins do not. Longer-chain polyphosphates cause the precipitation of both casein and whey proteins. These reactions are complex and not fully understood. Functions of phosphates in different types of dairy substitutes are summarized in Table 9 (see also FOOD ADDITIVES).

Table 9. Functions of Phosphates in Dairy Substitutes

Imitation dairy food	Phosphate type ^a	Function
milk-based pudding	SHMP, STP, TSPP, TSP	gelation
milk	DSP	heat stability
coffee whitener	DSP, DKP, Poly, SALP	pH control, reduce feathering, reduce oiling off
whipped topping	DSP, DKP, TSPP	pH control, whip stability, overrun improvement
sour cream	DSP, DKP, TSP, TSPP	improve body, reduce wheying
ice cream	DSP	protein stability
cheese	MSP, DSP, TSP, TSPP	modify protein, emulsifier

^a DKP = dipotassium phosphate; Poly = polyphosphate; and SALP = sodium aluminum phosphate. Other terms may be found in the text.

Composition and Processing

The composition of dairy substitutes is highly variable and generally represents the least-cost formulation consistent with consumer acceptance of the product. These imitations invariably have lower fat and protein levels than the dairy products that they are made to resemble. The gross compositions of filled milk, mellorine, synthetic milk, sour cream, coffee whiteners, whipped toppings, and cheese are listed in Table 10. A comparison of the composition of certain dairy products and their substitutes is presented in Table 11.

Table 10. Compositions of Whole and Filled Milk, Mellorines, and Imitation Dairy Products, wt %^a

Ingredient	Whole milk	Filled milk			Mallorines ^b			Nondai ry milk		Nondairy sour cream	Liquid coffee whiteners	Whipped toppings	Cassinate-based cheese
		A	B	C	A	B	C	A	B				
fat	3.8	3.	3.	3.				3.	3.				
		8	1	4				0	7				
protein	3.3	4.	3.	3.				1.	0.		1.0–3.0		

		1	3	1		0	9					
carbohydrates	4.6	5.	4.	5.		8.	6.					
		3	9	0		0	7					
minerals	0.7	0.	0.	0.		0.	0.					
		9	8	7		2	5					
vitamins A and D ^c	†		†			†	†					
riboflavin ^c	†	†	†	†		0	0					
water					62.5	62.0	61.1		82.0–84.0	80.0–85.0	46.0–75.0	
milk solids ^d					13.0	12.5	11.5		6.0–15.0			
sugar					14.0	13.0	12.0				6.0–15.0	
corn syrup solids					6.0 ^e	6.0 ^e	5.0 ^e			5.0–10.0		
vegetable fats					4.0	6.0	10.0		6.0–20.0	8.0–12.0	20.0–32.0	
stabilizer–emulsifier					0.5	0.5	0.4					
total solids					37.5	38.0	38.9					
stabilizers ^c								0.	†	0.1–1.0	0.02–0.5	0.1–1.5
								8				
emulsifiers ^c								0.	†	0.3–2.0	0.2–1.0	0.3–1.5
								4				
buffer ^c								0.	†		0.1–0.5	
								2				
sodium cassinate											1.0–5.0	15–35
calcium caseinate												10–35
oil (mp 32–43°C)												1–2
emulsifier salts												0–0.5
other proteins ^f												0–25
acid (adipic, lactic)												0.1–1.8

^a Refs. 8,9.

^b Also known as filled ice creams.

^c †, present; , absent.

^d Does not include fat.

^e Used 42 DE corn syrup. DE = dextrose equivalent.

^f May include gluten, soy, egg white, or gelatin.

Table 11. Comparison of Dairy Products and Substitutes^a

	Gross composition, g 100 g ^b				Minerals, mg 100 g ^b				Vitamins		
	Carbohydrates	Fat	Protein	Ash	Calcium	Phosphorus	Sodium	Potassium	A, IU	Riboflavin, µg	Total solids, g 100 g ^b
milk (fluid), whole	38.9	27.8	27.8	5.6	936	738	396	1142	111	1350	12.6
									2		
milk substitutes	52.8	31.7	7.4	4.0	25	248					12.1
liquid	61.8	27.7	6.8	4.5	136	673	636	3270			11.0
light cream	14.6	72.7	10.5	2.2	353	280	182	331	302	509	27.5
									0		
coffee whiteners	48.5	39.9	4.9	2.8	12	718	293	788	110	0	98.9
dry	49.1	36.5	5.0	2.7	16	625	258	768	110	0	98.9
	46.1	37.2	5.0	2.7	46	561	290	606	440	108	98.9
	48.7	35.8	4.9	3.0	12	62	146	1040	200	219	98.5
liquid	50.2	47.9	3.0	1.5	23	30	543	6	217	0	26.7
									0		
	42.0	48.0	5.0	2.0	70	155	245	300	250	0	20.0
	49.1	40.5	8.6	2.7	72	212	496	121	0	0	22.2
	37.9	52.6	2.6	1.1	52	57					21.1
heavy cream	7.8	85.3	5.6	1.2	190	149	98	134	351	268	41.0
									0		
whipped toppings	40.8	43.2	4.6	1.0	17	32	100	48	800	0	99.9
dry	40.6	45.4	5.7	0.6	12	46	97	6	137	0	98.9
									0		
liquid	25.7	58.5	0.0	0.5	15	3	198	13	104	0	39.3
									0		
	31.6	55.3	2.0	0.2	14	20	20	2	222	0	49.1
									0		
aerosols	22.0	58.0	10.7	2.3	310	276	255	310	254	145	35.5
									0		
dairy-based	20.4	63.3	8.5	2.3	274	267	413	209	265	189	41.2
									0		
nondairy	29.2	58.7	0.0	0.5	10	3	243	11	113	0	39.1
									0		
	23.9	67.7	7.6	0.5	23	71	173	6	137	0	39.3
									0		

^a Ref. 10.

^b Dry matter.

Processes used for the manufacture of dairy substitutes are essentially the same as for dairy product counterparts (1,5,11,12). The processes common to all fluid products include: ingredient blending, pumping, pasteurization or sterilization, homogenization, cooling, and packaging. Products can be manufactured by batch or by continuous processes. A large volume of imitation products are made in large plants by continuous processes. Ingredients may be either liquid or dry, or a combination of both. Liquid ingredients can be blended by the use of positive metering pumps to a mixing tank or by in-line blending. Dry ingredients must be hydrated during the blending operations. This is most commonly accomplished by the use of turbine mixers in processing vats or by incorporating the dry material through a high speed, centrifugal pump. The blending temperature depends upon the nature of the ingredients, but it must be above the melting point of the fat and sufficient to fully hydrate gums used as stabilizers. Pasteurization is generally carried out in high temperature short time (HTST) units, in which the homogenizer is integrated into the pasteurization system. Sterilization is either by retort or by UHT processing, followed by aseptic packaging (see also [FOOD PACKAGING; STERILIZATION TECHNIQUES](#)).

Margarine. Margarine or oleomargine, originally marketed as an imitation butter, has a recognized identity of its own. Fat blends typically used in margarine manufacture are listed in Table 12. The proportion of fat blend and other ingredients varies with the type of margarine and with the country of manufacture. Within a given country, variations in fat blends, aqueous phases, and other ingredients are fixed by legislation.

Table 12. Composition of Margarine Fat Blends, wt %^a

Fat	Margarine					
	A	B	C	D	E	F
palm kernel	32	37	36			55
coconut	37	36	40	50		
palm						15
hydrogenated cottonseed ^b				18		
groundnut	20	16			11	
hydrogenated groundnut	11	5			69 ^c	30 ^d
soybean		6	24	22	7	
sesame				10	13	
Total	100.0	100.0	100.0	100	100	100

^a Ref. 13.

^b Mp, 34°C.

^c Mp, 30°C.

^d Mp, 40°C.

Skim milk was initially used as the aqueous phase in margarine. Where the law allows, margarines may contain caseinates, whey proteins, or soy proteins as the proteins component in the aqueous phase. The addition to margarine of 0.01–0.1 wt % sodium caseinate in place of milk has been proposed to eliminate sticking during frying. Substituting soy proteins for milk would have the same effect.

In addition to traditional margarines, the following are also available: whipped margarine, margarine containing 15–40 vol % of finely and uniformly dispersed gas; liquid margarine, margarine that pours and is made by one of several methods that modify the fat system (9,10,12); and low fat margarine, margarine having a fat content below 80 wt % and generally 30–65 wt %, ie, imitation margarine or spread. By regulation margarine contains 80 wt % fat. A number of imitation margarines have been introduced that have fat contents of 35–65 wt % (16). In most cases, these products contain sodium caseinate, whey protein, or soybean proteins as the protein portion of the aqueous phase; emulsifiers; and sometimes gums. Emulsifiers of choice are generally mono- and diglycerides. Some spreads are made from a blend of vegetable fat and butter and butterfat. The addition of small amounts of butter to margarines has been reported to improve flavor.

Cheese. The evolution of imitation cheese has come from the substitution of milk components in the development of filled and nondairy cheese and development of a synthetic cheese based on the Chinese food sufu, a form of tofu, which is based on soybean curd.

Filled and Caseinate-Based Cheese. Filled cheese was made traditionally by adding vegetable fat to skim milk and homogenizing at low pressure to make a uniform emulsion, followed by traditional cheesemaking procedures. With the development of cheese processing, filled cheese was made from skim milk curd to which was added oil, flavoring, and phosphate and citrate emulsifying salts. The mixture was heated to about 55–66°C and the melted, processed, and filled cheese was then packaged. Skim milk curd was frequently treated with proteolytic enzymes to obtain a body similar to that obtained when processing aged cheese. The casein-based imitation cheeses are based on these processes. A variety of approaches have been suggested for making these products. One such product is made by the process shown in Figure 3 (14). The types of cheese that are commonly imitated are cheddar and mozzarella.

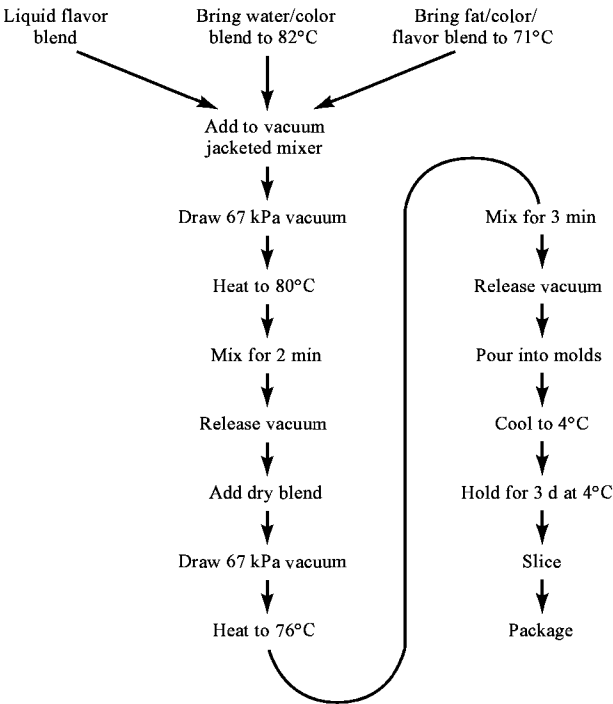


Fig. 3. Schematic process flow diagram for an imitation cheese product having the following formulation: dry ingredients, calcium caseinate (or rennet casein), 24.5 wt %; tapioca flour, 3.0 wt %; salt, 2.16 wt %; adipic acid, 0.6 wt %; vitamins and minerals, 0.1 wt %; sorbic acid (mold inhibitor), 0.5 wt %; fat–color blend, soybean oil hydrogenated to a Wiley melting point of 36°C, 21.3 wt %; lactylated monoglycerides, 0.05 wt %; red-orange coloring, 0.01 wt %; and flavoring blend, 0.23 wt % (15). To convert kPa to psi, multiply by 0.145.

Soybean-Based Cheese. In tofu cheese manufacture, the soybeans are cooked to give soy milk, then formed into a curd using calcium sulfate, and pressed to give tofu. The tofu is inoculated with *Actinomucor elegans*, fermented, salted, and aged to form sufu (soft cheese) (17). This product is widely used in the Orient and is gaining acceptance in the United States.

Interest in developing imitation synthetic cheese gained prominence in the middle 1960s. One such product is prepared from specific amounts of natural cheese, pregelatinized starch, a high protein binding agent (preferably soybean), water, and sugar (17). The mixture is heated to 90°C and extruded into the form of small strands. More recent approaches have been to combine caseinate and soy protein (enriched in the glycinin fraction) to make imitation processed cheeses. An imitation cheese spread based on soy protein and caseinate consists of margarine, 37.38 wt %; sodium caseinate, 5.58 wt %; maltrin dextrin, 5.50 wt %; modified food starch, 3.52 wt %; gelatin, 2.46 wt %; soy isolate, 2.28 wt %; whey, sweet, 1.66 wt %; acid blend, 0.86 wt %; lactic acid (50 wt %), 96.33; citric acid, dry (50 wt %), 3.60; acetic acid, cone (50 wt %), 0.07; emulsifier, 0.17 wt %; salt, 0.08 wt %; and water, 40.51 wt % (16).

Coffee Creams/Whiteners. Coffee-cream substitutes have been given the generic name of coffee whiteners and are available in liquid, frozen, and dry forms. The emulsifiers and proteins are key stabilizing factors with respect to the stability of the product in coffee. The type of emulsifier system varies with the type of coffee whitener. In addition, the method of processing has a significant effect on the stability and whitening power of the product. A typical process for a liquid coffee whitener would follow that given in Figure 4. Homogenization is a key process and the desired particle size is 0.2–1.0 micrometers. Rapid cooling after homogenization assists in producing a stable product. Both caseinate and soy protein-based products are marketed. Three typical formulas for whiteners are listed in Table 13 (18,19). There is a relationship between emulsifier and protein amounts with respect to the stability of a soy-based coffee whiteners, as shown in Table 14.

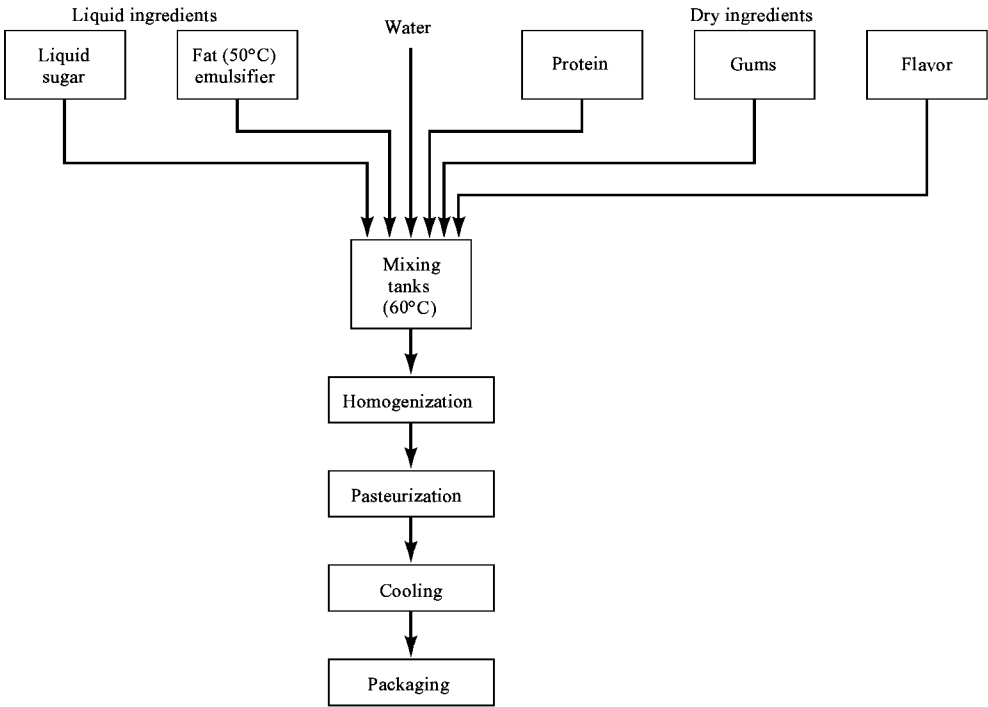


Fig. 4. General processing scheme for fluid imitation dairy products.

Table 13. Formulation of Caseinate and Soy-Based Coffee Whiteners^a

Ingredient ^b , wt %	Sodium caseinate-based	Soy protein-based	
		A	B
corn syrup solids	7.5	12.4	15.0
vegetable fat	10.0	9.5	10.1
protein	1.5	0.60	0.8
mono- and diglycerides	0.5		0.5
sodium-stearoyl-2-lactylate		0.1	0.2
sorbitan monostearate		0.2	
polysorbate 60		0.56	0.2
dipotassium phosphate	0.10	0.1	0.2

^a Refs. 18 and 19.
^b The primary ingredient, which makes up the remaining percentage, is water.

Table 14. Effect of Protein and Emulsifier on Coffee Whitener Stability^a

Sample	Concentrations of independent variables ^b				Height of feather, mm
	SSL ^c	ATMOS 150 ^d	TWEEN 60 ^d	Soy isolate	
A	0.10	0.35	0.56	0.70	0
B	0.10	0.20	0.32	0.60	coagulated

C	0.25	0.35	0.56	0.60	1.5
D	0.40	0.35	0.56	0.70	11
E	0.40	0.275	0.44	0.70	coagulated
F	0.40	0.275	0.44	0.60	27
G	0.25	0.275	0.44	0.80	42
H	0.25	0.275	0.44	0.60	23
I	0.25	0.35	0.56	0.70	18

^a Ref. 19.

^b The following concentrations in wt % were kept constant: hydrogenated coconut oil, 9.5; K₂HPO₄, 0.1; corn syrup, 12.4. Water is the primary ingredient, making up the remaining percentage.

^c Sodium stearoyl-2 lactylate.

^d Manufactured by Atlas Co.

Ice Cream and Frozen Desserts. Imitation frozen desserts include ice cream, sherbets, and specialty products, eg, milkshake bases. Filled ice cream, called mellorine in the United States, is the dominant product in the world market and is sold both as a soft-serve and as a hardened product. The composition and processing is as for ice cream, except for the substitution of vegetable fat for milk-fat. Frozen desserts based on caseinates or whey protein concentrates have not been widely marketed, because there has been no economic advantage to do so. Whey protein has been reported as being more functional than casein in ice cream manufacture (20). One formula for a whey–protein-based milkshake product in wt % is hydrogenated soybean oil, 4.0; sucrose, 11.0; salt, 0.02; dry whey–caseinate blend, 10.0; alginate, 0.26; carboxymethyl cellulose, 0.10; calcium lactate pentahydrate, 0.025; anhydrous citric acid, 0.1125; hydrolyzed cereal solids, 1.00; FD&C No. 3 (dry) red, 0.0005; imitation strawberry flavor (dry) 0.053; imitation strawberry flavor (liquid), 0.0150; and water, 73.414.

Soybean-based ice cream products, technologically feasible, are generally not in use because of flavor problems. An acceptable ice cream has been made by replacing 50% of the nonfat milk solids with a dried soy protein isolate made up of cheese whey (21). Chocolate flavor has been widely used to mask the flavor of soybean proteins in ice cream (see FLAVORS AND SPICES).

Milk. Imitation milks fall into three broad categories: filled products based on skim milk, buttermilk, whey, or combinations of these; synthetic milks based on soybean products; and toned milk based on the combination of soy or groundnut (peanut) protein with animal milk. Few caseinate-based products have been marketed (1,22,23). Milk is the one area where nutrition is of primary concern, especially in the diets of the young. Substitute milks are being made for human and animal markets. In the latter area, the emphasis is for products to serve as milk replacers for calves. The composition of milk and filled-milk products based on skim milk can be found in Table 10. Table 15 gives the composition of a whey buttermilk-solids-based calf-milk replacer, which contains carboxymethyl cellulose (CMC) for proper viscosity of the product.

Table 15. Dry Mass Composition of Milk-Substitute Formulation ^{a, b}

Component ^c , wt %	Protein base			Total 100%
	42% Buttermilk	30% Clotted whey (pH 1.5)	15% Sweet whey	
protein	13.9	3.9	2.0	19.8
lactose	17.6	19.2	11.0	47.9
mineral materials (ash)	3.4	1.8	1.4	7.2
lactic acid	5.0	4.8	0.6	10.4
Ca ²⁺	0.59	0.15	0.14	1.13
Mg ²⁺	0.09	0.03	0.05	0.17
K ⁺	0.76	0.03	0.36	1.15
Na ⁺	0.21	0.03	0.11	0.35
total cations	1.65	0.24	0.66	2.80
PO ₄ ³⁻	1.22	0.99	0.33	2.54
Cl	0.42	0.42	0.21	1.05
Ca P	1.48	0.48	1.30	1.36

^a Ref. 24.

^b Fat 10.5 wt % and CMC stabilizer 0.5% in all formulations, which are based on clotted-cream buttermilk.

^c Without vitamins and trace elements.

Guidelines for fluid milk substitutes (23) include recommendations regarding raw materials, methods of preparation, composition, sanitation, control of toxic substances, heat destruction of trypsin inhibitors, and recommendations for the distribution of filled or toned milk and products containing non-cow-milk components. Most of the cow-milk substitutes are based on soy bean or groundnut (peanut). Special considerations are given to protection from toxic materials that may be present in groundnuts and elimination of antinutritional factors in soybeans, eg, the trypsin inhibitor and hemagglutinin. Peanuts, in particular, and soybeans, to a lesser degree, may be contaminated with *Aspergillus* molds that form aflatoxins. The level of aflatoxin should not exceed 30 ppb on a dry basis in the protein product. Urease enzyme activity is used as an indirect measure of proper heat treatment to eliminate antinutritional factors.

Toned milk, containing peanut protein, was introduced in India in the 1960s under the name of Miltone. A typical process for toned milk is dispersion and hydration of protein in water at 60°C, addition of corn syrup followed by the addition of oil–emulsifier blend; addition of gums; two-stage homogenization, at 14 MPa (2000 psi), and then at 3.4 MPa (500 psi); pasteurization; cooling; and packaging (25).

Soy-Based Beverages. Soy milk is a traditional oriental product and has been the basis for development of dairy substitutes for western markets. Soybean-based beverages may be classified as (1): (1) traditional soy milks, unfermented, starting with uncommunited full-fat beans, full-fat or defatted soy flours, or soy protein concentrates; (2) traditional fermented yogurt-like products; or (3) simulated milks based on soy protein concentrates and isolates, ie, of the dairy type, fluid single strength, concentrated, or sterile, nonfat milk replacers such as dry blends of soy protein isolates with dry whey, infant-feeding beverages simulating human milk, and fermented yogurt-like products.

A process for making a soy milk having a minimum beany flavor is presented in Figure 5. The recommended compositions for a fluid and for a dry product are given in Table 16.

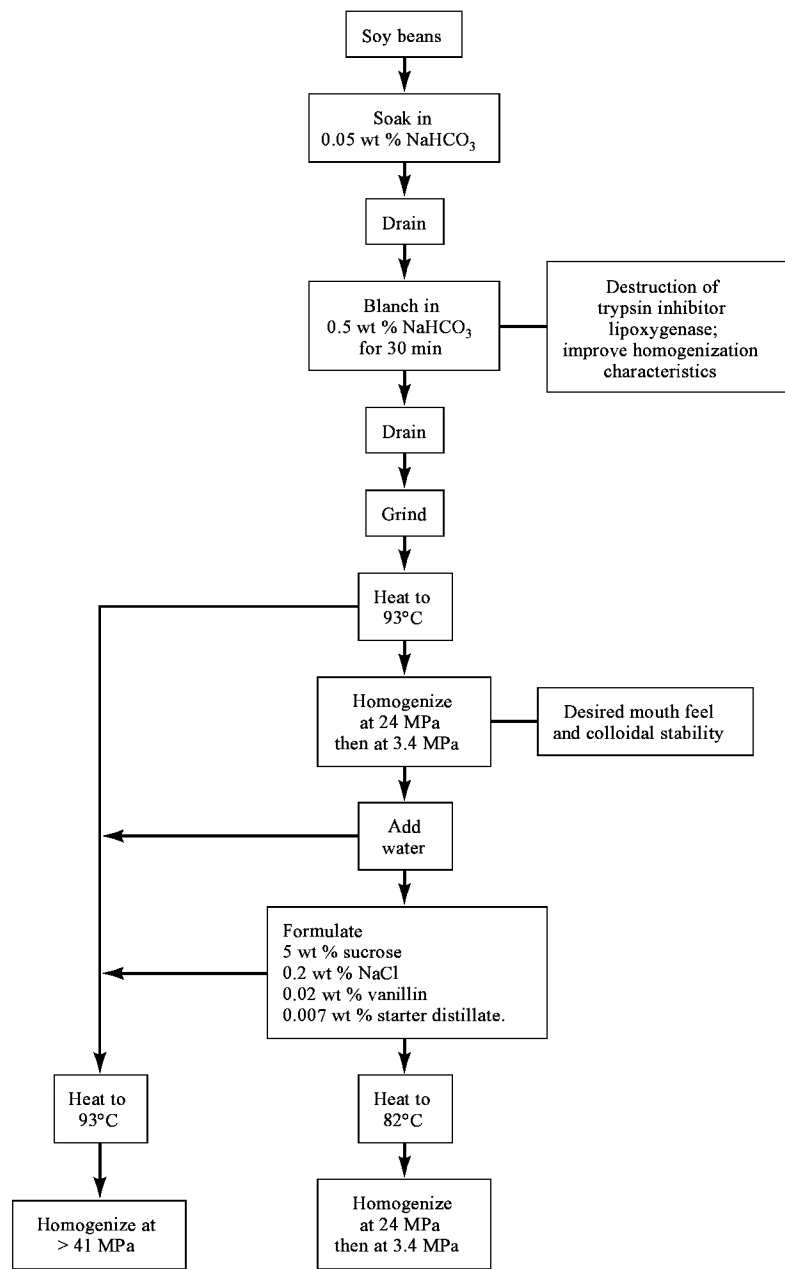


Fig. 5. Schematic process flow diagram for soy milk via the Illinois process (25). To convert MPa to psi, multiply by 145.

Table 16. Recommended Chemical Composition for Milk Substitute^a

Property	Fluid, sterilized	Dry
moisture, wt % max	89.0	3.5
fat, wt % min	2.0	18.0
nonfat solids, wt % min	9.0	
protein (N × 6.25), ^b wt % min	3.5	28.0
essential fatty-acid content, ^c wt % of fat	2.0	2.0
carbohydrates (sucrose or malt sugar–dextrin), wt % min	4.0	51.5
calcium, mg 100 mL ^d	100	800
vitamin A, µg 100 mL ^d	50	450
vitamin D, IU 100 mL ^d	40	360
iron, mg 100 mL ^d	0.4	4.0
thiamine, mg 100 mL ^d	0.15	1.0
riboflavin, mg 100 mL ^d	0.15	1.3
niacin,mg 100 mL ^d	1.0	8.0
folic acid, µg 100 mL ^d	5	50.0
pantothenic acid, µg 100 mL ^d	200	1600
vitamin B ₁₂ , µg 100 mL ^d	0.5	5.0
pyridoxine, µg 100 mL ^d	200	1600
ascorbic acid, mg 100 mL ^d	3.0	25.0
total ash, wt % max		7.0
acid-insoluble ash, wt % max		0.05

^a Ref. 29.
^b N total nitrogen.
^c Expressed as linoleic acid. Minimum value given.

^d Values given are minimum values. Units for dry material are per 100 grams.

Infant Formulas.

Milk-Based. Milk-based infant formula are sometimes referred to as humanized milk. These products are formulated to resemble the composition of human milk and contain a lower amount of carbohydrate than cow's milk (27). They are sold in three forms: liquid ready-to-use, liquid concentrates, and dry powders. A typical formulation is given in Table 17. A typical processing scheme involves adding skim milk to a mixing vat that is adjusted to 60°C; adding oil-emulsifier, minerals, and vitamins; pasteurizing; homogenizing at 21 MPa (3000 psi); standardizing composition; packaging; sterilizing; cooling; and storing.

Vegetable-Protein-Based. Most vegetable-protein-based infant formulas are soybean-based (28). Soybean-based formulas have served as useful replacements for milk in the nutritional management of infants who are allergic to cow's milk. Allergies to soybean milk also exist, so the substitution of soy milk for cow's milk does not always solve allergy problems. This concern is lessened by hydrolyzing the protein to reduce potential allergenicity.

Formulas tend to contain isolates as the protein source to eliminate or reduce the presence of carbohydrates that are the cause of flatulence and abnormal stools. Care is taken to provide adequate nutrition and to use proteins processed in such a way as to minimize or eliminate any antinutritional factors. The formulation of a typical soy-based infant formula is also given in Table 17.

Table 17. Commercial Infant Formulas

Components	A ^a	B ^a	C ^a	D ^b
protein	nonfat cow milk	demineralized whey and nonfat cow milk	nonfat cow milk and soy protein isolate	soy protein isolate
fat	vegetable oils	vegetable oils and oleo oil	corn oil	
added carbohydrates	lactose or corn syrup solids	lactose	sucrose	
	<i>Main constituents</i>			
protein, g 100 mL	1.5–1.6	1.5	3.6	2.0
fat, g 100 mL	3.6–3.7	3.6	1.6	3.6
carbohydrate, g 100 mL	7.0–7.1	7.2	6.6	6.8
minerals, g 100 mL	0.3–0.4	0.3	0.7	0.38
	<i>Caloric distribution</i>			
protein, 90	9	9	26	
fat, 90	48–50	48	27	
carbohydrate, 90	41–43	43	47	
	<i>Minerals</i>			
calcium, mg L	550–600	445	1000	700
phosphorus, mg L	440–455	330	800	500
sodium, meq L	11–17	7	17	300 ^c
potassium, meq L	16–28	14	32	710 ^c
chloride, meq L	12–14	10	29	530 ^c
magnesium, mg L	40–48	53	85	50
sulfur, mg L	130–160	145	310	
copper, mg L	0.4–0.6	0.4	1	0.5
zinc, mg L	2.0–4.2	3.2	4.0	5.0
iodine, µg L	40–69	69	100	150
iron, mg L	trace–1.5	12.7	18	12
	<i>Vitamins</i>			
vitamin A, RE L ^d	340–500	530	600	500
thiamine, µg L	400–650	710	750	40
riboflavin, µg L	600–1000	1060	900	600
niacin, mg L	7–8.5	7	10	9.0
pyridoxine, µg L	320–400	423	700	400
pantothenate, mg L	2.1–3.2	2.1	5	5.0 ^e
folacin, µg L	50–100	32	100	100
vitamin B ₁₂ , µ L	1.5–2.0	1.1	2.5	3.0
vitamin C, mg L	55	58	50	55
vitamin D, IU L	400–423	423	400	400
vitamin E, IU L	8.5–12.7	9.5	6.3	15

^a Ref. 27.
^b Ref. 28.
^c Units are mg L.
^d Vitamn A is reported as retinol equivalents per liter (RE = 5IU = 125 µg vitamin A or 3 µg carotene).
^e As mg L of pantothenic acid. Formula also contains 150 µg L each of biotin and vitamin K₁.

Whipping Creams and Whipped Toppings. Whipped topping is the term generally used for imitation whipping cream. Products are available as liquids to be whipped, frozen prewhipped products, and as liquids in aerosol cans that form whips upon release to the atmosphere (see AEROSOLS). Dry powdered products that are reconstituted and whipped are also available (18). Toppings were first made from nonfat dry milk or sodium caseinate as the protein source and vegetable fat as the source of fat. Most of these products are heavily dependent upon these constituents to impart the desired characteristics and to provide uniformity in the whipped topping. Generally, the toppings have a lower fat content than traditional whipping cream.

Isolated soy proteins have also been used in whipped toppings. Soy-protein-based toppings have a lower protein concentration than caseinate-based toppings. Formulations are adjusted to protein levels, and higher protein levels can result in off-flavors. Typical formulations for a liquid frozen, prewhipped product are given in Table 18.

Table 18. Formulations for Whipped Toppings^a

Ingredient	Liquid nondairy topping,	Frozen prewhipped topping, wt
	wt ^o %	^o %
water	55.25	53.40
palm-kernel oil	25.00	26.00
sucrose	18.00	18.00
isolated soy protein	0.40	1.20
sorbitan monostearate	0.30	0.35
flavor	0.30	0.30
lecithin	0.20	
mono- and diglycerides	0.20	
polysorbate 60	0.20	0.30
disodium phosphate	0.15	0.15
guar gum		0.30

^a Ref. 18.

Miscellaneous. Miscellaneous products having market significance include sour cream, chip dips, milkshake bases, puddings, yogurt and fat-reduced forms of all substitute dairy foods. These products have been formulated from both caseinates and soybean proteins.

Economic Aspects

Substitutes exist for most dairy products. In the United States, commercial dry mixes or bases without fat that are available include: filled buttermilks, filled cheese, filled coffee creams, filled dip products, filled eggnog, filled frozen desserts, filled milk, filled sour cream, filled whipping cream, imitation ice cream, imitation milk and milk products, and imitation coffee cream. Complete dry mixes available include imitation coffee cream, imitation flavored dips, imitation frozen desserts, imitation half-and-half, imitation milk and milk drinks, imitation sour cream, and imitation whipping cream. In addition, filled and imitation products are sold under generic names such as chip dips, coffee whiteners, margarine, spreads, and whipped toppings. These dairy substitutes have identities apart from dairy products. Many are sold under fanciful or trade names, for example, coffee whiteners are sold under a wide range of trade names that include: Coffee Light, Coffee Mate, Coffee King, Complement, Cremora, and Crodamix. Most of these substitute products, including margarine, often contain a milk component.

Of the filled and imitation products, margarine has made the greatest penetration into the sale of dairy foods. The market share of butter, margarine and spreads in different countries is shown in Table 19. In Great Britain up to 50^o % of the butter and creams sales have been replaced by margarine. The amount of replacement is greater in the U.S., reaching about 82^o % in 1987. However, U.S. per capita consumption of butter increased between 1977 and 1987. Indeed, spreads and blends of butter and margarine have increased sales (2).

Table 19. World Market Share of Butter, Margarine, and Spreads^a

Country	Butter, ^o %	Margarine, ^o %	Spreads ^b , ^o %
Australia	34	60	2
Finland	50	35	10
Japan	18	80	2
United Kingdom	29	57	14
United States	16	82	2
Sweden	15	57	28

^a Ref. 29.

^b Spreads are products with fat reduced below 80^o % required by standards of identity for butter and margarine.

Other dairy substitutes have penetrated the U.S. market to the extent of 1^o % for fluid whole milk, 58^o % for creams, <1% for low-fat milk, 6–7^o % for cheeses, 10^o % for evaporated and condensed milks, and 2^o % for ice cream (30). About 60^o % of the substitute and imitation cheese sold in the United States is being used as the cheese material for pizza (2).

In the Philippines, the sale of filled milk had become 85^o % of the dairy products market by the early 1970s, reflecting convenience as a purchase incentive rather than price. Filled condensed and evaporated milk has a market share of 10, 27, 54, 69, and 77^o % in the Netherlands, Mexico, Malaysia, Phillippines, and Thailand respectively (30). Imitation cream has an 8^o % market share in the United States, 11^o % in Spain, and 33^o % in Canada (30). Areas in the world expected to show the greatest growth in the sale of substitute and imitation dairy products are Canada, Ireland, and Mexico (30).

Based on brand name products, the total number of branded substitute and imitation dairy products worldwide is estimated to exceed 1000. Almost all multinational food companies market one or more dairy substitutes.

Cost is a significant factor in the consumer's acceptance of substitute dairy foods. Table 20 shows the relative cost of substitute fat and protein in the various substitute foods. A comparison of retail prices of selected dairy products and corresponding substitutes in four supermarkets in the midwestern United States in the Fall of 1992 are shown in Table 21. In all cases the prices for the substitutes are lower than the prices of the respective dairy product. The smallest price margin is in the area where the substitute products are advertised as fat-reduced or cholesterol-free.

Table 20. Relative Composition and Ingredient Costs of Dairy Products and Substitutes^a

Substitute ^b	Component relative to dairy product, ^o %		Cost relative to dairy product, ^o %	
	Fat	Protein	A	B ^c
milk-filled	86	100	70	72
coffee whitener	56	80	27	40
whipped topping	78	100	33	59
ice cream	100	86	52	55
cheese	76	89	70	81

^a Ref. 10.

^b Products other than filled milk based on caseinate.

^c Corrected to same fat and protein level as dairy product.

Table 21. Prices of Dairy Products and Substitutes

Category	Price, \$ 100 g	
	Dairy Food	Dairy Substitute
coffee cream		
liquid	0.35	0.15–0.28
frozen	0.35	0.24–0.32
dried	1.23 ^a	0.35–0.76
whipped toppings		
aerosol	0.95	0.79–0.87
frozen	0.48 ^b	0.31–0.66
sour cream	0.44	0.26–0.34
butter	0.35	
margarine		0.07–0.19
lite spreads		0.29–0.33
spreads ^c	0.35	0.31–0.33
shredded cheese	1.05 ^d	0.57–0.83

^a Dried whole milk.
^b Heavy cream (nonfrozen).
^c Spreads citing the word butter on the label.
^d Cheddar.

Based on ingredient cost comparisons, the fat cost in dairy products is 40–90^o   of the total ingredient costs. In contrast, fat makes up 16–30^o   of the ingredient cost of substitute products. This is related to both the higher price of milk fat and the higher amount of fat that is generally in dairy products as compared to substitutes. The cost of milk fat has been declining, however. But as milk fat price goes down, milk protein prices increase, shifting the emphasis from milk fat substitution to milk protein substitution.

Regulatory Aspects

The legal status of dairy substitutes varies widely among countries. In the United States, the FDA has held that filled products should be nutritionally equivalent to the products they resemble. In the case of imitation milks, the FDA proposed regulations for nutritional equivalency (31). A section of the Food, Drug and Cosmetic Act defines misbranded foods, and the FDA has set up standards of identity for foods under this part of the law (32) which includes standards for imitation milks, cheese, and creams (32).
The definition of margarine varies from country to country (33). The U.S. Code of regulations on margarine is given in Ref. 33 and on imitation ice cream, called mellorine, in Ref. 34. Regulations on food labeling, proposed in 1991, have not as of this writing been promulgated.

BIBLIOGRAPHY

"Synthetic and Imitation Dairy Products," in *ECT* 3rd ed., Vol. 22, pp. 465–498, by W. James Harper, Ohio State University.

1. A. H. M. van Gennip, *International Dairy Federation Document 107*, International Dairy Federation, Brussels, Belgium, 1978, pp. 92–93.
2. *International Dairy Federation Document 239*, International Dairy Federation, Brussels, Belgium, 1987, pp. 1–20.
3. T. E. Furia, *Handbook of Food Additives*, 2nd ed., CRC Press, Boca Raton, Fla., 1980.
4. K. Sabharwal and D. G. Vakaleris, *J. Dairy Sci.* **55**, 277, 283 (1972).
5. F. Winklemann, *Imitation Milk and Milk Products*, Food and Agricultural Organization of the United Nations, New York, 1974, pp. 1–117.
6. J. Morita, *Jpn. J. Dairy Food Sci.* **29**(1), 26 (1980).
7. M. E. Matthews, *N.Z. J. Dairy Sci. Technol.* **13**(3), 149 (1978).
8. Economic Research Service, *Market Research Report No. 947*, U.S. Dept. of Agriculture, Washington, D.C., 1972.
9. *U.S. Dairy Coun. Dig.* **39**(2), 7 (1968).
10. H. H. Moede, *USDA Economic Research Service, MTS-177*, U.S. Dept. of Agriculture, Washington, D.C., 1979, pp. 19–23.
11. T. J. Hendrick, *Dairy Ind.* **34**(3), 127 (1969).
12. W. J. Harper and C. W. Hall, *Dairy Technology and Engineering*, Avi Press, Newport, Conn., 1976.
13. M. K. Switzer, *Margarine and Other Food Fats*, Leonard Hill, Ltd., London, 1956, pp. 240–273.
14. **U.S. Pat. 4,104,413** (Aug. 1, 1978), J. D. Wynn (to Anderson, Clayton & Co.).
15. **Brit. Pat. 1,517,956** (July 19, 1978), J. D. Wynn (to Anderson, Clayton & Co.).
16. R. F. Bennet and J. A. Ohren, *Food Prod. Dev.* **8**(5), 15 (1974).
17. J. J. Jonas, *J. Milk Food Technol.* **39**(1), 39 (1975).
18. C. W. Kolar, I. C. Cho, and W. L. Watrous, *J. Am. Oil Chem. Soc.* **56**, 389 (1979).
19. R. J. Pearce and W. J. Harper, *J. Food Sci.* **47**, 680 (1982).
20. P. A. Huse, C. Towler, and W. J. Harper, *Dairy Sci. Technol.* **19**(3), 255–261 (1984).
21. M. D. Wilding, *J. Am. Oil Chem. Soc.* **56**, 392 (1979).
22. **Brit. Pat. 1,564,800** (Apr. 16, 1980), J. Bodor (to Unilever, Ltd.).
23. *P.A.G. Statement of Milk Substitutes, No. 6*, Protein Advisory Group of the United Nations System, New York, 1970.
24. **U.S. Pat. 3,433,643** (Mar. 18, 1969), P. P. Noznich and C. W. Tatter (to Beatrice Foods).
25. *P.A.G. Guideline No. 15 on Milk Substitutes*, Protein Advisory Group of the United Nations System, New York, June 1972.
26. *Quick Frozen Foods*, **34**(9), 100 (1972).
27. S. J. Foman, *Infant Nutrition*, W. B. Saunders Co., Philadelphia, Pa., 1974, pp. 377–382.
28. W. A. B. Thomson, *J. Am. Oil Chem. Soc.* **56**, 386 (1979).
29. G. Agren and A. Oklund, *Bulletin* 11(ix), Protein Advisory Group (P116) of the United Nations, New York, 1972, pp. 19–23.
30. *International Dairy Federation Bulletin* 270, International Dairy Federation, Brussels, Belgium, 1992.
31. *Fed. Reg.* **38**, 20748 (Aug. 2, 1973).
32. Food, Drug, and Cosmetic Act, Section 403.
33. M. T. Gillies, *Shortenings, Margarines, and Food Oils*, Noyes Data Corp., Park Ridge, N.J., 1974, pp. 219–220.
34. *21 FCR 120*, Chapt. 1, Part 45.
35. *21 FCR 120*, Chapt. 8, Part 20.

W. James Harper
Ohio State University

DAMMAR, DAMAR.

See RESINS, NATURAL.

DATABASES

Data collection and database statistics have been published periodically since the 1970s with the onset of the worldwide database industry (1). Available information includes the growth of the database industry as represented by the increase in the number of database records, online searches, existing databases, database entries in the computer readable databases (CRDB), and database producers and vendors. Databases are analyzed in terms of geographic region and status of the producers, where the latter refers to the sector of society from which they come, eg, government, commerce industry (for-profit), or not-for-profit (NFP), which includes academia and mixed sectors. Databases are also considered in terms of the form or representation of the data and intellectual content in the database and the medium for access and or distribution (see General References).

The significance of databases to chemical technology arises because there are so many publications in active research areas, that it is nearly impossible for an individual to read or recall all citations appearing in journal articles, conference proceedings, dissertations, technical reports, statistical tables, directories, or other documents such as those in the intellectual property arena. Thus a number of database organizations, also called electronic libraries or knowledge banks, have developed. These provide rapid access to the international literature for researchers who have access to computers, modems, and a telecommunication network, as well as a knowledge of a straightforward search and retrieval command language or a menu-driven system (2).

Chemically related database searches can be used to establish concepts and patentable ideas. For instance, searches have identified researchers using particular monomers in a potentially patentable latex formulation; found precedents for a polymeric emulsifier; summarized publications of people being considered as consultants, expert witnesses, employees or speakers to an industrial group; and provided market description information for a new pigment manufacturing firm to identify target markets.

Some driving forces for undertaking a search are competitive and strategic factors related to market needs. Information from databases might help determine the markets for an emerging technology; how amenable the technology is to production scale-up; the patents held by competitors; what regulatory issues apply; what toxicity data are available and what is known about safe-handling or shipping; who manufactures the chemicals required, and what the physical and chemical properties are; what the manufacturing costs and pricing implications are; where company headquarters, subsidiaries, and executive information are located; and what information can be retrieved from molecular structure searches.

Growth in Databases

Size of the database industry can be measured in terms of the number of database records, databases, database entries in CRDB, database producers, database vendors, or online searches. Growth in number of database entries, databases, producers, and vendors is plotted in Figure 1. The slowest growth is at the vendor level, the fastest growth is in database files.

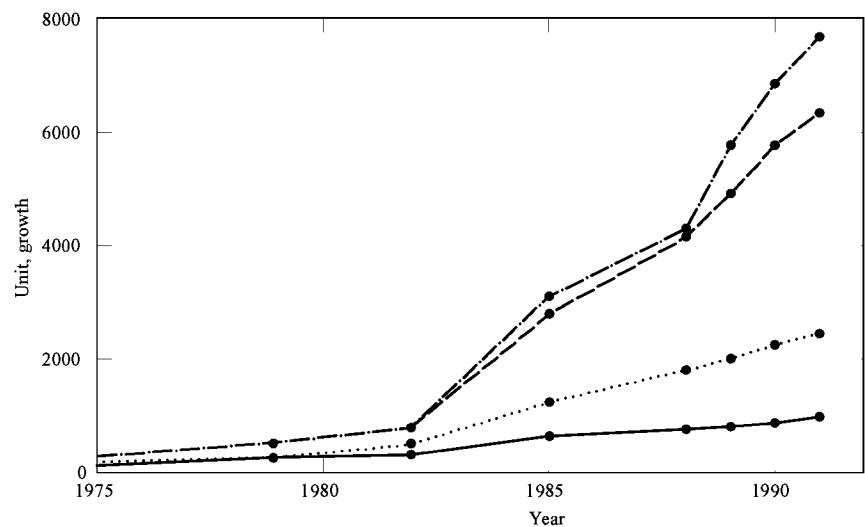


Fig. 1. Growth in number of (—), database vendors; (· · ·), producers; (---), entries; and (- · -), databases.

The 1992 edition of the CRDB (1) covers 7637 databases and unique subfiles in 6261 entries. Some of these entries represent families of databases, ie, sets of subfiles, rather than single databases. Databases are sometimes called files and members of families of databases are sometimes called subfiles. Some 848 of the files are "obit" entries, meaning that the producer no longer maintains the file, the producer no longer makes the file publicly available, the file is no longer available from the vendor indicated in a prior edition of CRDB, or the status of the file and its producer could not be verified. The statistics herein relate to the 6261 database entries known in 1991, rather than to the individual databases and subfiles.

Many database producers provide online access to their databases or distribute their databases on compact disc read-only-memory (CD-ROM) and so are also considered vendors or producer vendors. Whereas numerical growth in vendors is indicated (Fig. 1), the success of the database industry is largely a result of the transition of the information industry from paper-based to computer-based services (see COMPUTER TECHNOLOGY; INFORMATION RETRIEVAL, INFORMATION STORAGE MATERIALS). Thus industry growth can also be measured in terms of the increase in use of computer-readable databases as exemplified by the number of searches.

Database Records. From 1975 through 1991, database records increased from 52 million to 4.06 billion, a factor of 77, and the number of databases grew from 301 to 7637, a factor of 24. Database entries grew from 301 to 6261, a factor of 20. The number of producers continues to grow somewhat more slowly in part because individual producers create multiple databases. There were 2372 producers in 1991 and the average producer

published 3.2 databases. The number of database vendors, who offer services from numerous databases, increased from 105 (1975) to 933 (1991). The growth in number of database records has not been proportionate to the growth in number of databases. That is, the average size of a database has not remained constant. The growth in number of records was rather slow until 1983 when it reached 310 million. From 1983 to 1984 it more than tripled; from 1984 to 1987 it doubled; and from 1987 to 1991 it nearly doubled again, reaching more than four billion records. The average database, which contained 173,000 records in 1975, reached approximately 500,000 records in 1985. The average database size is highly skewed, however, because 259 databases contain more than one million records. In fact, of the 259 have more than 100 million records, and another 45 have between 10 million and 100 million records. Excluding these 259 very large databases, in 1991 the average database entry contained 70,645 records and the average database contained 59,950 records.

The entities counted as database records vary widely but generally range from 200 to 2000 words. Records may be citations, abstracts, news stories, biographical records, unique chemical substance property data, recipes, time series, software programs, images, or descriptions or listings of virtually anything. Perhaps the single largest factor affecting the average database size was the increase in commercially available business databases that started in the mid-1980s. Among the very large databases, and in descending order of frequency, are databases containing bibliographic information (references, abstract, or full text); company information; time series; patent information; telephone and address listings; personnel, biography, and employment listings; chemical data; demographic data; news; vocabulary (terms, words, or thesauri) entries; trademarks; and procurements and contracts.

Online Searches. The real success or acceptance of computer-readable resources and the technology for accessing them is indicated by the growth in the use of the databases. Use can be measured both in terms of the number of searches, which translates to connect time, and in terms of expenditures for searching. Some data regarding the use of databases through primary vendors of word-oriented databases in the United States are available (3). These vendors include Mead Data Central, West Publishing Company, DIALOG Information Services, ORBIT Search Service, BRS Information Technologies, and the U.S. National Library of Medicine. Figure 2 shows the growth in online database use for databases on principal U.S. systems offering word-oriented databases. Searches increased from 750,000 in 1974 to over 34 million in 1991. These figures represent a specific subset of vendors, however, and if transactional, eg, stocks, electronic ordering, credit checking, airline reservations, etc, and consumer service systems were added, the number of searches would be much larger.

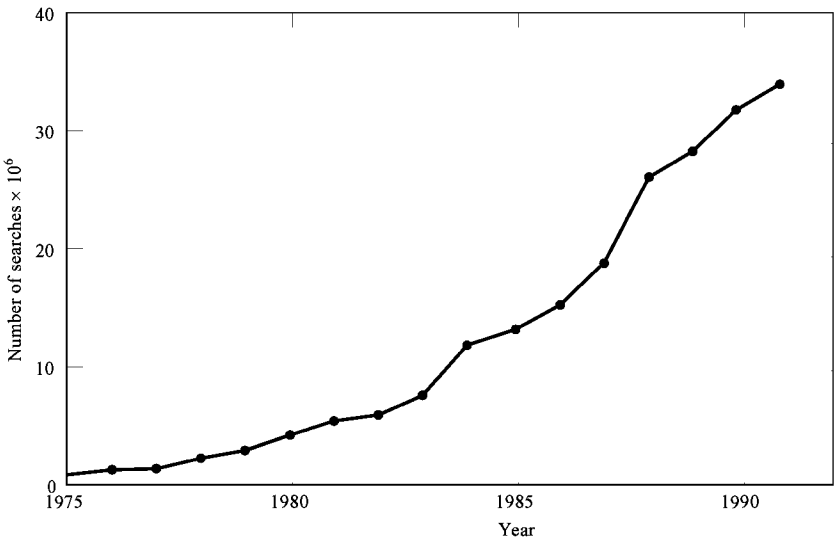


Fig. 2. Growth in number of online database searches on principal U.S. vendors of word-oriented databases.

Whereas Figure 1 represents worldwide data for all types of publicly available databases, Figure 2, showing growth in database searches, represents only a portion of those databases, ie, U.S. usage of word-oriented databases in the information center library market.

Future Growth. The database industry is dynamic, changing and growing every year. There is no sign of leveling-off in the 1990s. New media for distribution and access to databases have increased the potential for attracting users. In particular, the development and distribution of CD-ROM databases has greatly increased the use of computer-readable databases in universities and colleges where cost has often been a barrier to access. At the same time, there has been a decrease in the use of online searching of those databases that are also distributed on CD-ROM. Decreases have been noticed particularly on ERIC, PsycINFO, NTIS, and MEDLINE, which are all databases well used in academia.

Classification of Databases

Form of Data. Databases can be classified in many ways. One method is by form of data representation, ie, data may be in the form of words, numbers, images, or sounds. The corresponding databases may then be considered to be word-oriented, number-oriented, image-oriented (video), or sound-oriented (audio). Data representation affects file structures and software for search and data retrieval. Thus the structures and search techniques vary considerably among these four basic classes. Table 1 gives databases as classified by form of data representation.

Table 1. Database Classes Normalized to One Class per Database

Class	Number ^{a,b}									
	1988		1989		1990		1991		1992 ^c	
word-oriented	2797	(69)	3370	(70)	4080	(72)	4491	(72)	4925	(70)
number-oriented	1136	(28)	1236	(26)	1298	(23)	1370	(22)	1533	(22)
image	14	(<1)	34	(<1)	113	(2)	145	(2)	272	(4)
audio	1	(<1)	2	(<1)	16	(<1)	27	(<1)	83	(2)
electronic services	90	(2)	134	(3)	170	(3)	172	(3)	146	(2)
software	4	(<1)	10	(<1)	12	(<1)	55	(1)	39	(<1)
Total	4042	(100)	4786	(100)	5689	(100)	6261	(100)	6998	(100)

^a Value is number of database entries in CRDB.
^b Value in parentheses is percentage.
^c Estimates for individual databases in 1992 can be calculated by multiplying the percentages by 7907.

For purposes of tracking statistics on database classes, several subclasses for word-oriented and number-oriented databases have been established. Word-oriented databases can be subdivided into: bibliographic, patent trademark, directory, dictionary, full-text, and other categories. The first publicly

available databases were word-oriented. In the 1960s and 1970s, most databases were bibliographic and thus fell within the word-oriented class. Numeric databases can be subdivided into: transactional, statistical, time series, properties, and other categories. Publicly accessible image and audio databases did not exist until the mid-1980s. Audio and image databases are being developed for Apple Macintosh II, IBM and IBM clones, and NEXT microcomputers. With the popularization of hypertext and CD-ROM systems, more multimedia databases are expected. Multimedia CD-ROM interweaves audio, video, and ASCII text so that these can be played simultaneously and can even use different audio tracks for different languages.

Many databases can be classified in multiple ways because of multiple type data, eg, text and numeric data, text and image data, image and audio data, etc. Also included in the data presented in Table 1 are two additional classes of databases, electronic services and software. Both of these data types could also be classed by form of representation because of use of words and numbers. However, the way in which these databases are used is different and they have special characteristics. Thus they are presented as additional classes. Whereas electronic information services such as bulletin boards, electronic mail, and electronic conferencing contain data that are transitory and nonarchival, these must be included among databases because several of the principal vendors sell these services in the same way as database search services are sold.

Electronic mail and electronic conferencing services are not counted as separate databases, but when such electronic services are included with a database they are counted as aspects of the associated database and the database is classified accordingly. Several commercial online services make software databases available online, eg, the Microsoft Programmers Library, MacTutor, and Atari RoundTable. Several producers offer CD-ROM databases containing software. Such software databases contain the coding and documentation for software packages and so are classified as software databases. These are different from directories of software that are classed as directory databases.

Databases in CRDB have been coded according to the six classes described (1). Table 1 provides the number and percentage of databases associated with the six database classes. Despite multiple classes assigned to some databases, all classes are normalized to the 6998 database entries in CRDB. Table 2 presents the same data without the normalization. Because a single database entry may have more than one class assignment, there are more database entry-class assignments than database entries in CRDB.

Table 2. Database Classes Multiple Classes per Database Entry

Class	Number ^{a,b}									
	1988		1989		1990		1991		1992	
word-oriented	3147	(69)	3409	(70)	4213	(72)	4661	(72)	6497	(71)
number-oriented	1278	(28)	1250	(26)	1360	(23)	1422	(22)	1974	(21)
image	16	(<1)	34	(<1)	98	(2)	151	(2)	358	(4)
audio	1	(<1)	2	(<1)	16	(<1)	28	(<1)	109	(1)
electronic services	101	(2)	136	(3)	178	(3)	179	(3)	193	(2)
software	5	(<1)	10	(<1)	12	(<1)	57	(1)	52	(1)
Total	4548	(100)	4741	(100)	5877	(100)	6498	(100)	9183	(100)

^a Value is number of database entry-class assignments in CRDB.

^b Number in parentheses is percentage.

The percentage of word-oriented databases continues to increase at a faster pace than numeric databases. The number of image databases was 358 in 1992, greater than twentyfold increase from 1989. Audio databases rose from 1 in 1988 to 109 in 1992.A breakdown of the subclasses within word-oriented databases is given in Table 3, showing that full-text databases have surpassed bibliographic ones. Directory databases are the third most numerous.

Table 3. Word-Oriented Databases Subclasses

Subclasses	Number ^{a,b}									
	1985 ^c		1989		1990		1991		1992	
bibliographic	1094	(57)	1223	(36)	1367	(32)	1425	(31)	1715	(26)
patent trademark ^d			58	(2)	80	(2)	85	(2)	47	(<1)
full-text	535	(28)	1412	(41)	1786	(42)	2040	(44)	3077	(47)
directory	287	(15)	707	(21)	952	(23)	1074	(23)	1611	(25)
dictionary	10	(<1)	9	(<1)	23	(1)	32	(<1)	47	(<1)
other					4	(<1)	5	(<1)	0	(0)
Total	1926	(100)	3409	(100)	4212	(100)	4661	(100)	6497	(100)

^a Value is number of database entries in CRDB.

^b Value in parentheses is percentage.

^c In 1985 each database was classified in only one way; in subsequent years databases were classified in one or more ways.

^d In 1985 patent trademark information was included as part of bibliographic.

Geographical Origin. The 6261 database entries listed in CRDB are produced in eight geographic regions: Africa, Asia, Australia, the Far East, Eastern Europe, Western Europe (including the United Kingdom), North America, and South America. No databases from Antarctica are listed. The geographical distributions of database entries and databases are shown in Table 4. Asian databases are from China (both Mainland and Taiwan), Hong Kong, India, Israel, Philippines, Japan, Singapore, Thailand, and the CIS. The only regions having more than 1000 databases are North America (>4400) and Western Europe (>1400). The majority of databases, 4368 in 1991, are held in the United States. Figure 3 compares U.S. and non-U.S. database entries. Counties, other than the United States, having 100 or more databases are England, Canada, Germany, France, Japan, and Australia. Databases are produced in a variety of languages.

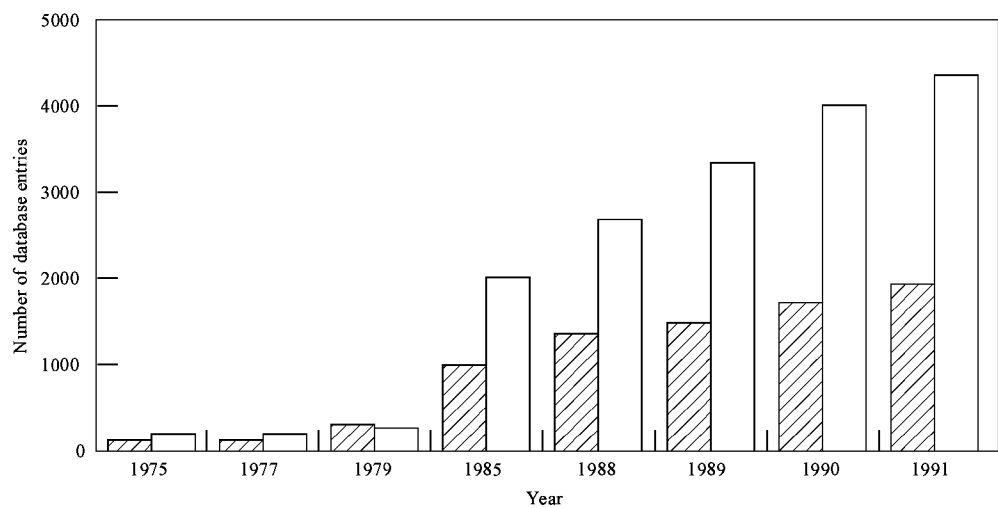


Fig. 3. U.S., □, versus non-U.S., ▨, database entries. Growth by year.

Table 4. 1991 Database Entries and Databases by Geographic Region

Region	Database entries ^a		Databases ^{a,b}	
Africa	7	(<1)	8	(<1)
Asia	28	(<1)	34	(<1)
Australia	119	(2)	146	(2)
Far East	155	(3)	189	(3)
East Europe	11	(<1)	13	(<1)
West Europe	1473	(24)	1797	(24)
North America	4424	(71)	5396	(71)
South America	44	(<1)	54	(<1)
Total	6261	(100)	7637	(100)

^a Value given in parentheses is percentage.
^b Normalized from database entries reported.

Subject Categories. The determinant for user selection of a database is usually subject matter. That is, when chemical information is desired, a chemical database is selected. The form or media of the database is of secondary importance. The type of search may dictate the need for a full-text or statistical database. If none exists, however, a bibliographic database in the topic area may be used to locate full-text or numeric compilations in hard-copy form.

Databases can be categorized by subject as given in Table 5, where the numbers reflect subject category assignments rather than numbers of databases. That is, many databases can be categorized by two or more subjects eg, the Bio-Business database can be categorized as belonging to both the health life sciences and the business categories. Any database reflecting a reasonable percentage of a subject category was assigned a code to that category. A subject category that was incidental to the database was not assigned.

Table 5. Databases by Subject

Category ^a	Number ^b							
	1989		1990		1991		1992	
business	1687	(33)	1956	(33)	2101	(33)	2624	(33)
general	327	(6)	416	(7)	450	(7)	700	(9)
health life science	576	(11)	651	(11)	690	(11)	728	(9)
humanities	184	(3)	216	(4)	248	(4)	314	(4)
law	447	(9)	531	(9)	574	(9)	885	(11)
multidisciplinary ^c	335	(7)	368	(6)	366	(6)	296	(4)
news	186	(4)	233	(4)	291	(4)	385	(5)
social sciences	393	(8)	418	(7)	453	(7)	447	(6)
sci tech engineering	996	(19)	1154	(19)	1210	(19)	1492	(19)
Total	5131	(100)	5943	(100)	6383	(100)	7841	(100)

^b Values in parantheses are percentages.
^a Multiple subject categories may be assigned to a database.
^c Academic.

Business databases remain number one among all database categories, followed by science technology engineering and then by health life science. Many business databases are cross-coded with law and or with news. Through 1988 there was also a consumer interests category referring to databases covering information of interest or use to consumers. The consumer interests category was distinct from databases about consumers. Most databases that are about consumers relate to market studies or demographics and are categorized as business or social sciences.

Medium of Distribution. Another way of looking at databases is in terms of the recording medium used for distribution or access. Table 6 indicates the number and percentage of databases recorded on various types of media. In 1991, there were 8159 instances of database media combinations for the 7636 databases in 6261 database entries. In Table 6, the distribution of instances over the types of media is used to normalize media for the number of entries and databases.

Table 6. Media for Database Distribution or Access

Medium ^a	1990			1991		
	Number of occurrences	DBs ^b	Entries	Number of occurrences	DBs ^b	Entries
online	4018	3608	3041	4170	3903	3200
batch	1252	1124	948	1321	1236	1014
CD-ROM	715	642	541	1019	954	782
diskette	626	562	474	695	651	533
magnetic tape	906	814	686	954	893	732
<i>Total</i>	<i>7517</i>	<i>6750</i>	<i>5690</i>	<i>8159</i>	<i>7637</i>	<i>6261</i>

^a A database may be available on several media.
^b DBs Databases.

Producers. The producers of databases are sometimes called database publishers because they make public their databases. Some producers publish hardcopy counterparts to databases and so are publishers in the traditional sense; others publish data only in electronic form. Database producers are responsible both for the determination of content and for database production. Most producers offer their databases for lease or license to private organizations or database vendors. Vendors offer database search services to the marketplace on a fee basis. An increasing number of producer vendors such as Mead Data Central, U.S. National Library of Medicine, and DRI McGraw-Hill (formerly Data Resources), offer search services (batch or online) from their own databases as well as from the databases of other products.

Producers of databases can be classified in terms of a country's infrastructure. Identifiable groups are the government, commerce industry, not-for-profit (NFP) which includes academe, and mixed status. In the 1960s and 1970s, most databases were produced by U.S. governmental organizations such as the National Aeronautics and Space Administration (NASA) and Atomic Energy Commission (now the U.S. Department of Energy). In the late 1960s and 1970s, the mostly professional society-based NFP databases also gained prominence as reflected in increasing use. Both government and NFP databases continue to be important resources, particularly in the sciences. As outlined in Table 7, however, between 1977 and 1991, the governmental and NFP databases decreased as a percentage of total available databases, whereas commercial databases have continued to increase, comprising 70% of the total databases in 1991.

Table 7. Databases Classified by Producer Status, %

Producer status	1985	1988	1989	1990	1991	1992
government	21	20	17	17	16	15
commerce industry	57	65	68	68	70	75
not-for-profit ^a	11	13	12	12	12	9
mixed	11	2	3	3	2	1
<i>Total</i>	<i>100</i>	<i>100</i>	<i>100</i>	<i>100</i>	<i>100</i>	<i>100</i>

^a Includes academe.

Although government databases have decreased percentagewise, many commercial databases are built on government data. Census data, for example, are collected at considerable governmental expense. Commercial database producers can take these data, add value to them, and then resell this information as a commercial database product.

Some databases are produced either by organizations having mixed status or by more than one organization. The majority of these databases is produced by a mixture of governmental organizations or an organization representing more than one government, such as the United Nations or the European Economic Community.

Vendors. Database vendors provide value added processing of databases and offer search services, online and or batch, or distribute CD-ROM products to database users. Vendors also provide the technology for accessing databases.

Value is added to databases in several ways: through database preparation (loading); by means of the special capabilities of the vendor's search software; and through related services such as online document ordering and selective dissemination of information (SDI) current awareness. During the 1960s, database vendors in the United States were often referred to as tape spinners. In Europe, vendors are generally called hosts. Vendor is the preferred term among database users and it is also the term most frequently used in the literature to designate the organizations that offer computer-based database search services.

Among the primary vendors of abstracting and indexing databases are DIALOG Information Services, BRS Information Technologies, ORBIT Search Service, U.S. National Library of Medicine, Questel, Pergamon Financial Data Services, and Data-Star. Principal vendors of full-text databases are Mead Data Central, West Publishing Co., NewsNet Inc., Dialog Information Services, Inc., DataTimes Corp., and STN International. Vendors of numerica business databases include the WEFA Group (formerly Wharton Electronic Forecasting Associates and Chase Econometrics), Reuters Information Services (Canada) Ltd. (formerly I.P. Sharp Associates), and DRI McGraw-Hill. Examples of vendors of numeric scientific databases are CAN SND and FIZ Karlsruhe's INKADAT. Vendors of consumer oriented databases include CompuServe Information Service and PRODIGY.

Producers who serve as vendors for their own databases are more numerous than vendors that offer services from databases produced by other organizations. More than 650 producer vendors plus some 270 traditional commercial database vendors are listed in CRDB(1). Only those vendors that offer search services or distribute CD-ROMs for databases other than their own come under this vendor classification (1). Vendors that offer services solely from databases they themselves produce are listed as database producers.

Connecting to Databases

Online Connections. One highly reliable telecommunications network is DIALNET, offered by DIALOG. DIALNET offers dial-up access from over 50 principal U.S. cities at 300, 1200, 2400, and higher baud rates to 9600. Other networks for users in the United States and Canada include SprintNet or TYMNET networks. INWATS and INTERNET access is also available. In other countries, timeshare telecommunications access is available through the local PTT; rates vary from country to country. Access is also available through services such as TWX, international telex, outward WATS or FTS; for these services, the user bears the communication charges directly (4).

Using a Personal Computer. A personal computer can be used to communicate with online services when the following three items are available (5): (1) some form of communications software which must be synchronous and compatible with the particular computer. DIALOGLINK, available from DIALOG, may be used with IBM or IBM-compatible personal computers. A shareware package called PROCOMM has been reviewed favorably for lab computer usage (6). A number of similar software packages are also available for Apple computers; (2) a modem which must also be compatible with the computer. Modems may be internal or external to the computer, but if external, a serial interface card must be internally installed; and (3) a dedicated telephone line which allows no transfers or call-waiting. Call-waiting options may be disabled by computer command in dial-out code in some areas.

Using a Terminal. Dial-up terminals may also be used to access a vendor's databases; however, the terminal, personal computer, word processor, or microcomputer must all be compatible with specific functionalities. For example in order to access the DIALOG service these must all be TTY compatible; have ASCII coding; and be compatible for asynchronous transmission in full-duplex mode for communication through DIALNET,

TELENET, or TYMNET at 30, 120, 240, or 960 character per second transmission speed. In addition, one also needs an EIA compatible (within the United States) or CCITT-compatible (non-U.S.) modem for dial-up.

The telecommunications protocol includes one start bit, two data bits, one even parity bit, and one stop bit, for a total of 10 bits per character. Alternatively, one start bit, eight data bits, no parity, and one stop bit (10 bits character) may be used. The start bit, part of most protocols, may not be discussed in the documentation.

Databases Available

Databases may be viewed as a research tool for those in highly competitive fields such as science, engineering, and technology. This tool makes information available so that its user can make decisions based on comprehensive and timely data (7). A guide to commercially available databases relating to topics associated with science and technology follows.

Searching. A truncation feature (?) that allows word variation, eg, "mass spectrometry or mass spectroscopy" is used. Title searching is accomplished by using the added modifier " ti" to bring up only titles. Commands to retrieve information generally use a protocol such as type-set number format choice number of records.

Chemistry. An alphabetical listing of the chemical databases available in 1992 together with brief descriptions follows (8): *AGRICOLA*, also called Agricultural Online Access, formerly called Cataloging and Indexing (CAIN), *Agricultural Library Forum (ALF)*, all aspects of agricultural sciences, U.S. National Agricultural Library products and services; *Agrochemicals Handbook*, active chemical substances used in pesticides, plant growth regulators, pest repellents, and synergists; *Analytical Abstracts Online (AA)*, general analytical chemistry and techniques; *Applied Science and Technology Index*; *Beilstein Current Facts in Chemistry*, organic compound structures and substructures; *Beilstein Online; Biological and Agricultural Index*; *Chemical Abstracts (CA) File*; *CA Registry File*, substance information and identification; *CA Search*, formed through the merger of two previously existing files, CA Condensates (CACon) and CA Subject Index Alert (CASIA); *Cambridge Structural Database (CSD)*, crystal structures of organic and organometallic compounds analyzed by x-ray or neutron diffraction methods; *CAOLD*, all areas of chemistry and chemical engineering; *CApreviews*, all areas of chemistry and chemical engineering; *CASREACT*, chemical catalysts, products, reactants, reagents, and solvents, covering general organic chemistry; *Chapman and Hall Chemical Database*, formerly called Heilbron, organic and organometallic chemical compounds, natural products, pharmaceuticals and antibiotics, bulk industrial chemicals, synthetic reagents and laboratory solvents, compounds having unusual structural, physical, or chemical properties and compounds of the metalloid elements; *CHEMEST*, physical and chemical properties of organic compounds; *Chemical Abstracts Service Source Index (CASSI)*, publications monitored for input to CAS products; *Chemical Business NewsBase (CBNB)*, the chemical industry, with emphasis on European activities; *Chemical Journal of the Association of Official Analytical Chemists (CJAOAC)*, analytical chemistry; *Chemical Journals of John Wiley & Sons (CJWILEY)*, polymer and bipolymer science; *Chemical Journals of the American Chemical Society (CJACS)*, chemistry, biochemistry, chemical engineering, and related topics; *Chemical Journals of the Royal Society of Chemistry (CJRSC)*, chemistry; *Chemical Journals of VCH Verlagsgesellschaft (CJVCH)*, chemistry research; *Chemical Reactions Documentation Service (CRDS)*, new synthetic methods in organic chemistry; *Conference Papers Index (CPI)*, conference papers in science and technology; *Current Awareness in Biological Sciences (CABS)*, biological sciences and related areas; *Current Contents Search*, alternatively called *CC Search*, very broad physical and life sciences coverage; *C13 Nuclear Magnetic Resonance Database*, ¹³C nmr spectra of organic compounds; *DIPPR Data Compilation of Pure Compound Properties*, 39 properties for more than 1000 chemical compounds; *Ei CHEMDISC*, chemical engineering; *Electronics Material Technology News*, hightech specialty chemicals and materials used in electronic products; *EMTOX*, toxicology; *Environmental Fate Data Bases*, fate of organic chemicals; *Epidemiology Information System (EIS)*, food contaminants and their effects on human health; *European Chemical News (ECN)*, European chemical industry; *European Directory of Agrochemical Products*; *FDA Drug Bulletin*; *FOODS ADLIBRA*, the food industry; *General Science Index*; *Gmelin Formula Index (GFI)*, inorganic and organometallic chemistry, HODOC, organic compounds and spectral data; *Infrared Search System (IRSS)*; *Infrared Spectral Database*, infrared spectra of organic compounds; *Inorganic Crystal Structure Data Base (ICSD)*, alternatively called *Gmelin/Flz Inorganic Crystal Structure Data Base*; *Japan Technology*, science and technology; *Japanese Information on Scientific and Technical Topics (JAPINFO)*; *JICST File on Science and Technology*, international scientific and technological literature; *JICST File on Science, Technology and Medicine in Japan*; *Kirk-Othmer Encyclopedia of Chemical Technology Online*; *Log P Database*, measured and calculated partition coefficients for organic compounds; *Marine Environmental Data Information (MEDI)*, oceanographical observations; *Microbiology Abstracts, Section A: Industrial and Applied Microbiology*; *Microbiology Abstracts, Section C: Algology, Mycology and Protozoology*; *National Union Catalog Codes*, libraries cited in the *Chemical Abstracts Service Source Index*; *NIST Crystal Data File*, alternatively called *Cristaldon*, formerly called NBS Crystal Data Identification File; *ORBCHEM*, chemical nomenclature; *PAPERCHEM*, scientific and technical aspects of the pulp and paper industries; *PASCAL*, alternatively called, *Programme Applique a la Selection et a la Compilation Automatiques de la Litterature* (PASCAL), science and technology; *Pest Control Literature Documentation (PESTDOC)*, formerly called Pesticidal Literature Documentation; *Pharmaceutical and Healthcare Industries News Database (PHIND)*; *Physical Property Data Service (PPDS)*; *PHYTOTOX*, alternatively called Plant Toxicity Data; *Plant Growth Regulator Abstracts*; *Pollution Abstracts*; *Radiation Chemistry Data Center Database*; *Registry of Mass Spectral Data*; *Review of Agricultural Entomology*, formerly called Review of Applied Entomology, Series A: Agricultural; *Review of Medical and Veterinary Mycology*; *SCISCAN*, natural, physical, earth, environmental, biomedical, and life sciences; *SciSearch*, natural, physical, earth, environmental, biomedical, and life sciences; *Seed Abstracts*; *Standard Pesticide File (SPF)*; *Structure and Nomenclature Search System (SANSS)*, chemical substances, including compounds, covered by databases made available online through CIS; *The Chemical Monitor*, chemical instrumentation; *THERMALLOY*, thermodynamic properties of metallic alloys; *THERMODOC*, thermodynamic properties of inorganic materials; *THERMOCOMP*, thermodynamic properties of inorganic elements and compounds, covering enthalpy of formation, entropy, specific heat, and other details; *TRC Vapor Pressure Datafile*, alternatively called *TRC Thermophysical Property Datafile 1: Vapor Pressure*; *Who's Who in Technology*, and *World Patents Index (WPI)*.

Biochemistry and Biophysics. An alphabetical listing of biochemistry and biophysics databases (8) available as of 1992 follows: *Biochemistry Abstracts, Part 1: Biological Membranes*, formerly called Biological Membranes Abstracts; *Biochemistry Abstracts, Part 2: Nucleic Acids*, formerly called Nucleic Acid Abstracts; *Biochemistry Abstracts, Part 3: Amino Acids, Peptides, and Proteins*, formerly called Amino Acids, Peptides, and Protein Abstracts; *BioCommerce Abstracts and Directory*; *Biological and Agricultural Index*, *BIOSIS Previews*, life sciences dealing with animals (including humans), plants, and microorganisms; *Chemical Abstracts' (CA) File*; *CA Registry File*; *CA Search*, formed through the merger of two previous files, CA Condensates (CACon) and CA Subject Index Alert (CASIA); *Calcified Tissue Abstracts*, all aspects of calcium and related mineral metabolism; *CAOLD*, all areas of chemistry and chemical engineering; *CApreviews*, all areas of chemistry and chemical engineering; *CASREACT*, chemical catalysts, products, reactants, reagents, and solvents; *Chemical Journals of John Wiley & Sons (CJWILEY)*, polymer and bipolymer science; *Chemical Journals of the American Chemical Society (CJACS)*; *Chemical Journals of the Royal Society of Chemistry (CJRSC)*; *Chemoreception Abstracts*, the sciences of taste, smell, and related phenomena; *Conference Papers Index (CPI)*, conference papers in science and technology; *CSA Life Sciences Collection*, formerly called IRL Life Sciences Collection; *Current Awareness in Biological Sciences (CABS)*; *Current Biotechnology Abstracts (CBA)*; *Current Contents Search*, alternatively called *CC Search*; *EMBASE plus*, alternatively named *Excerpta Medica*, biomedicine, human medicine, and related disciplines; *INSPEC*, formerly called Information Services for the Physics and Engineering Communities; *JICST File on Science, Technology, and Medicine in Japan*; *Log P Database*, measured and calculated partition coefficients for organic compounds; *Microbiology Abstracts, Section B: Bacteriology*; *Nutrition Abstracts and Reviews, Series B: Livestock Feeds and Feeding*; *Physics Abstracts (PA)*; *PHYTOTOX*, alternative name: *Plant Toxicity Data*, and *Who's Who in Technology*.

Geochemistry and Mineralogy. An alphabetical listing of geochemistry and mineralogy databases (8) available as of 1992 follows: *American Gem Market System (AGMS)*, the gem and jewelry industry; *Applied Science and Technology Index*; *Aquatic Sciences and Fisheries Abstracts (ASFA)*; *Australian Earth Sciences Information System (AESIS)*; *Chemical Abstracts' (CA) Registry File*; *Conference Papers Index (CPI)*, conference papers in science and technology; *Current Contents Search*, alternatively called *CC Search*; *Examine*, the earth sciences; *General Science Index*; *GeoArchive*; *Geobanque*, formerly called Banque de Donnees du Soussol; *GEOBASE*, formerly called GEOABS; *Geo Abstracts*; *GEOLINE*, the geosciences; *Geological Reference File (GeoRef)*; *GEOPAC*, Australian earth sciences; *GEOS*, geosciences; *Geoscience Data Index for Alberta (GEODIAL)*, all aspects of Alberta Geology; *IMAGE*, economic and exploration geology, mining engineering, extraction geology; *International Research and Evaluation-Information and Technology Transfer Database (IRE-ITTD)*, all aspects of science and technology; *JICST File on Science and Technology*, international scientific and technological literature covering such topics as Japanese and foreign chemistry and chemical industries; *JICST File on Science, Technology, and Medicine in Japan*, Japanese scientific and technological literature covering such topics as the chemical industry; *MARINELINE*, marine research and technology; *McGraw-Hill CD-ROM Science and Technical Reference Set*, science and technology, including physical, earth, and life sciences, and engineering; *Michigan Natural Features Inventory Program (MNFI)*, special animals and plants, natural ecological communities, and geological and other special features of Michigan's natural heritage; *Oceanic Abstracts (OA)*; *PASCAL-GEODE*, alternatively called *Geode*, earth sciences; *Petroleum Abstracts*, alternative name: *TULSA Data Base*, petroleum exploration, production, and development; *Publications of the U.S. Geological Survey*, earth sciences; *Rock Analysis Storage System (RASS)*, field geochemistry and petrology; *SCISCAN*, natural, physical, earth, environmental, biomedical, and life sciences; *SciSearch*, natural, physical, earth, environmental, biomedical, and life sciences; *Sea Grant Network (SGNET)*, marine science; *Soviet Science*

and Technology (SST); *Stimline*, formerly part of Geoline, industrial minerals and mineral resources; *Who's Who in Technology*; and *Yukon Bibliography* (YKB).

Biology and Medicine. An alphabetical listing of biology and medicine databases (8) available as of 1992 follows: *Advanced Medical Information Services*, Japanese health and medical industry; *AgBiotech News and Information*; *Agrar Forschungsvorhaben*, agricultural research projects; *AGRICOLA*, alternative name: *Agricultural Online Access*, formerly called Cataloging and Indexing (CAIN); *Agricultural Library Forum (ALF)*, all aspects of agricultural Sciences; *AIDSLINE*, aspects of Acquired Immune Deficiency Syndrome (AIDS); *Allied and Alternative Medicine*; *Animal Breeding Abstracts*; *Animal Disease Occurrence*; *Aquaria and Fish Forum*; *Aquatic Information Retrieval (AQUIRE)*; *Aquatic Sciences and Fisheries Abstracts (ASFA)*; *Arzte Zeitung Datenbank*, medicine and health-related sciences; *Australian Marine Research in Progress*; *Avline*, alternative name: *Audiovisuals Online*, audiovisual material in clinical medicine; *BioBusiness*; *BioExpress*; *Biological and Agricultural Index*; *Biology Digest*; *BIOSIS Previews*; *BMA Press Cuttings Database (BMAP)*, medical news and related ethical and sociological issues; *CAB Abstracts*, agricultural science; *CAT-LINE*, alternative name: *Catalog Online*, books and serials in the biomedical sciences; *Chemical Hazards Response Information System (CHRIS)*, transport of hazardous chemicals; *CISTI Monographs*, formerly called CISTI Current Catalogue; *CONF*, science technology, engineering, medicine, and health sciences; *CISTI Serials*, serials in all areas of science, technology, and medicine; *Colleague Mail*, formerly called MEDLINK, medical and health care topics; *Collective Catalogue of Belgium (CCB)*; *Company Directory Database*, formerly called Pharmacontacts, pharmaceutical, agrochemicals, and animal health organizations; *Comprehensive Core Medical Library (CCML)*, formerly called Critical Care Medicine Library; *Conference Papers Index (CPI)*, conference papers in science and technology; *CSA Life Sciences Collection*, formerly called IRL Life Sciences Collection; *Current Awareness in Biological Sciences (CABS)*; *Current Contents Search*, alternative name: *CC Search*; *DHSS-MEDTEH*, alternative name: *Medical Toxicology and Environmental Health*; *DIGS*, alternative name: *Domestic Information on the General Subjects*, Korean Periodicals Index; *Directory of Federally Supported Research in Canadian Universities*, formerly named Information Exchange Centre for Federally Supported Research in Universities, federally supported university research in Canada; *DIRLINE*, alternative name: *Directory of Information Resources Online*, information on health and biomedical organizations; *EMBASE plus*, alternative name: *Excerpta Medica*, biomedicine, human medicine, and related disciplines; *Emergindex System*, emergency and critical care medicine; *EMFORENSIC*, forensic science; *EMHEALTH*, public health, social medicine, and hygiene; *Excerpta Medica CD: Cardiology*; *Excerpta Medica CD: Drugs and Pharmacology*; *Excerpta Medica CD: Gastroenterology*; *Excerpta Medica CD: Immunology and AIDS*; *Excerpta Medica CD: Neurosciences*; *Excerpta Medica Vocabulary (EVOC)*; *ExpertNet*, medical malpractice and personal injury expert witnesses; *Forensic Science Database (FORS)*; *Forthcoming Events*, meetings, seminars, workshops, and special events in the life sciences; *General Practitioner*, general practice medicine; *General Science Index*; *Health Periodicals Database*; *Helminthological Abstracts*, former name: Helminthological Abstracts, Series A: Animal and Human Helminthology; *Index Veterinarius*; *Indice Medico Espanol (IME)*, alternative name: *Spanish Medical Index*, clinical and experimental medicine; *Information System for Hazardous Organics in Water (ISHOW)*; *International Information System for the Agricultural Sciences and Technology (AGRIS)*; *International Medical Tribune Syndicate*; *International Research and Evaluation-Information and Technology Transfer Database (IRE-ITTD)*; *Japanese Information on Scientific and Technical Topics (JAPINFO)*; *JICST File on Medical Science in Japan*; *JICST File on Science and Technology*; *JICST File on Science, Technology, and Medicine in Japan*; *Journal Watch*; *LEXIS Medical Malpractice Library*; *Maggie Mae's PET-NET and Co.*, animals and pets, animal health care, animal activities, animal preservation; *Marine Biotechnology Abstracts*; *Marine Environmental Data Information (MED)*, oceanographical observations; *McGraw-Hill Publications Online*, former name: McGraw-Hill Business Background; *MEDIC*, Finnish medical literature; *Medical and Psychological Previews (PREV)*; *Medical Subject Headings Vocabulary File*, alternative name: MeSH Vocabulary File, medical vocabulary thesaurus terms; *MEDIS*, medicine; *MEDLARS Name Authority File*, personal names, corporate names, and classification decisions; *MEDLINE*, alternative name: *MEDLARS Online*; *NIOSH Technical Information Center Database (NIOSHTIC)*, all aspects of occupational safety and health; *Nordisk Samkatalog for Seriella Medicinska Publikationer (Nordser)*, biomedical journals available in the Nordic countries; *NCR Publications*, National Research Council of Canada publications; *Nursing and Allied Health Database*; *Oceanic Abstracts (OA)*; *Oceanographic Literature Review*; *PASCAL*, alternatively called *Programme Applique a la Selection et a la Compilation Automatiques de la Litterature*, science and technology; *Permanent Inventory of Agricultural Research Projects in the European Community (AGREP)*; *Pharmaceutical and Healthcare Industries News Database (PHIND)*; *Pig News and Information*; *Plantas Medicinales (PLAMED)*, alternative name: *Herbal Remedies*; *Postgraduate Medicine*, general medicine; *Protozoological Abstracts*; *PTS Newsletter Database*; *Readers' Guide Abstracts*, topics of general interest; *Readers' Guide to Periodical Literature*; *Review of Medical and Veterinary Entomology*, formerly called Review of Applied Entomology, Series B: Medical and Veterinary; *Review of Medical and Veterinary Mycology*; *Right-To-Know Planning Report*, right-to-know rules, regulations, and policies; *Scientific American*, applied physical, life, and social sciences; *Scientific American Medicine (SAM)*, current developments in the 15 subspecialties of clinical medicine; *SCISCAN*, natural, physical, earth, environmental, biomedical, and life sciences; *SciSearch*, natural, physical, earth, environmental, biomedical, and life sciences; *Sea Grant Network (SGNET)*, marine science; *SEIBT*, German industrial companies; *SERLINE*, alternative name: *Serials Online*, biomedical serials; *Small Animals*, former name: Small Animal Abstracts; *Social Sciences Index*; *SOMED*, alternative name: *Sozialmedizin*; *Sportwissenschaftliche Forschungsprojekte (SPOFOR)*, research projects involving sports, including physical education; *Swedish Drug Information System (SWEDIS)*, drugs available in Sweden; *Swemed*, all areas of medicine; *The Medical Letter on Drugs and Therapeutics*; *The Medical RoundTable*, medical and health topics; *The New England Journal of Medicine*, all areas of medicine; *The Physician and Sportsmedicine*, sportsmedicine; *Union List of Scientific Serials in Canadian Libraries*; *Veterinary Bulletin*; *Veterinary Literature Documentation (VETDOC)*; *Who's Who in Technology*; and *Zoological Record Online*, alternative name: *ZR Online*.

Engineering, Mathematics, and Physics. An alphabetic listing of engineering, mathematics, and physics databases (8) available as of 1992 follows: *AgBiotech News and Information*, agricultural biotechnology; *Agrar Forschungsvorhaben*, agricultural research projects; *AGRICOLA*, alternative name: *Agricultural Online Access*, formerly called Cataloging and Indexing (CAIN); *Agricultural Engineering Abstracts*; *Agricultural Library Forum (ALF)*; *Agriculture Victoria-Library Catalogue (AVID)*; *Agroforestry Abstracts*; *Applied Science and Technology Index*; *Asian Geotechnology Engineering Database (AGE)*, alternative name: *AGE Database*, geotechnical engineering; *Australian Engineering Database (ENGINE)*, Australian engineering literature; *Australian Marine Research in Progress*, Australian research projects in marine sciences; *Automotive Information and News Service*, automotive industry in Western Europe, the United States, and Japan; *Banque de Donnees Internationales de Biometrie Humaine et d'Ergonomie (ERGODATA)*, alternative name: *International Data Base of Human Biometry and Ergonomics*; *BEFO*, managerial aspects of engineering; *BIIPAM-CTIF Data Base*, alternative name: *Banque de L'Information Industrielle Pont-a-Mousson*, metallurgy, foundry techniques, the iron industry, corrosion, coatings, materials resistance; *Biological and Agricultural Index*; *BNA Daily News*; *BODIL*, building and housing; *Business Periodicals Index*; *C-CORE Database*, cold ocean engineering; *Chemical Abstracts (CA) FILE*; *Ca Search*, formed through the merger of two previous files, CA Condensates (CACon) and CA Subject Index Alert (CASIA); *CAOLD*, all areas of chemistry and chemical engineering; *CAprreviews*, all areas of chemistry and chemical engineering; *CASE Strategies*, computer-aided systems engineering; *CERCLIS Database of Hazardous Waste Sites*; *CETIM Database*, mechanical engineering industry; *Chemical Engineering*, chemical process industries; *Chemical Engineering and Biotechnology Abstracts (CEBA)*, plant and process chemical engineering, covering both theoretical and practical aspects; *Chemical Journals of the American Chemical Society (CJACS)*; *Chemical Journals of the Royal Society of Chemistry (CJRSC)*; *Chemical Safety NewsBase (CSNB)*, health and safety in chemical and allied industries; *Chemical Week*, chemical process industry; *Chickpeas and Pigeonpeas*, chickpeas and pigeonpeas, plant physiology, nitrogen fixation, storage and quality, nutrition and utilization, handling and processing, and economics; *CIS Abstracts*, alternative name: *CISILO*, occupational health and safety; *Civil Engineering DataBase (CEDB)*; *CMP Publications Electronics File*, electronic engineering design, industry, and purchasing news; *Cold Regions Data Base*, all subjects relating to Antarctica, the Arctic, and subantarctic islands; *COMPENDEX*, alternative name: *Computerized Engineering Index*, all fields of engineering; *COMPENDEX PLUS*, all fields of engineering; *Computer and Mathematics Search*; *Computer and Information Systems Abstracts*, computer science in the broadest sense; *Conference Papers Index (CPI)*, conference papers in science and technology; *Conferences in Energy, Physics, and Mathematics*; *Cross Section Information Storage and Retrieval System (CSISRS)*, neutron, photon, and charged particle cross sections; *CSA Engineering*; *Current Contents Search*, alternative name: *CC Search*; *Current Research in Britain*; *Current Technology Index (CTI)*, former name: British Technology Index (BTI); *Dairy Science Abstracts*; *DECHEMA Environmental Technology Equipment Databank (DETEQ)*, manufacturers and suppliers of plants, apparatus, equipment, and machinery; *DECHEMA Research Institutes Databank (DERES)*, research and teaching establishments active in chemical engineering and biotechnology; *DECHEMA Thermophysical Property Data Bank (DETERM)*; *DESY-HEPI*, alternative name: *German Information System for High Energy Physics*; *DIGS*, alternative name: *Domestic Information on the General Subjects*, Korean Periodicals Index; *DMS/FI Contractors*, U.S. and international defense aerospace programs actively monitored in the DMS Market Intelligence Reports services; *DMS/FI Market Intelligence Reports*, former name: DMS Online, U.S. and world defense and aerospace markets; *Dokumentation Kraftfahrzeugen Database (DKF)*, motor vehicle design, construction, and manufacturing; *Dokumentation Maschinenbau Database (DOMA)*, alternative name: *Literaturdatenbank Maschinenbau*, mechanical engineering; *DTIC Manpower and Training Research Information System (MATRIS)*; *Ei CHEMDISC*, chemical engineering; *Ei ENGINEERING MEETINGS*, all fields of engineering; *Ei Page One*, engineering literature; *Electrical and Electronics Abstracts (EEA)*, all areas of electronics; *Electronic Engineering Times*, news and analysis of, as well as job openings in, the electronics industry; *Electronics and Communications Abstracts*; *ENERGIE*, formerly called Data Compilations in Energy (ECOMP), energy research and technology literature on fossil fuels; *ENERGIRAP*, alternative name: *ENERGI*, high energy and nuclear physics; *ENGINEERING AND INDUSTRIAL SOFTWARE DIRECTORY (EISD)*; *ENR: Engineering News-Record*, engineering and construction topics; *Environmental Fate (ENVTROFATE)*, fate or behavior of chemicals which are released into the environment; *FLUIDEX*, fluid engineering; *Fusion Power Report (FPR)*, fusion energy; *Gaz-Physique-Orsay Database (GAPHYOR)*, properties of atoms and molecules and their interactions, and the macroscopic properties of gases and plasmas; *General Science Index*; *Geomechanics Abstracts*; *Government-Industry Data Exchange*

Program (GIDEP), former name: Interagency Data Exchange Program (IDEP), qualification and evaluation test data on parts, components, and materials from tests performed by industry and government activities; *Groundnuts; Highway Research Information Service (HRIS)*; *IBSEDEX*, formerly called International Building Services Index, mechanical and electrical services associated with all types of buildings; *ICC Key Notes Market Research*; *ICONDA*, alternative name: *CIB International Construction Database*, IHS International Standards and Specifications; Industrial Environment; Information on Roads (INROADS), formerly called Australian Road Research Documentation (ARRD); *Innovator's Digest (ID)*, former name: Information for Innovators; *INSPEC*, former name: Information Services for the Physics and Engineering Communities; *International Nuclear Information System (INIS)*, peaceful applications of nuclear science and technology; *International Research and Evaluation-Information and Technology Transfer Database (IRE-ITTD)*, all aspects of science and technology; *International Road Research Documentation (IRRD)*; *ISIS Software Infobank*, commercial, technical, and scientific applications software products available in Austria, Switzerland, and Germany; *ISMEC: Mechanical Engineering Abstracts*, former name: Information Service in Mechanical Engineering; *Japan High Tech Review*, Japanese high technology industries; *Japan Technology*, science and technology; *Japanese Information on Scientific and Technical Topics (JAPINFO)*, science and technology; *JICST File on Science and Technology*, international scientific and technological literature covering such topics as Japanese and foreign chemistry and chemical industries; *JICST File on Science, Technology, and Medicine in Japan*; *Kirk-Othmer Encyclopedia of Chemical Technology Online*, chemical technology; *KSEA*, alternative name: *Korean Scientists and Engineers Abroad*; *LIDAS*, motor vehicles and automotive engineering; *Livres Bibliotheque Sackay Database (LJBISAC)*, books and conference proceedings pertaining to all aspects of nuclear energy and physics; *Marine Environmental Data Information (MEDI)*, oceanographical observations; *Mathematics Abstracts (MATH)*, former name: Mathematics and Related Subjects Data Base; *MathSci*, former name: *MATHFILE*; *McGraw-Hill CD-ROM Science and Technical Reference Set*; *Mechanical Engineering*; *MERLIN-TECH*, electrical and electronic engineering; *Military Robotics Sourcebook*; *Mobile Data Report*, mobile data communications industry, *MOVE*, *SAE Global Engineering Technology*, self-propelled vehicle technology; *MPD Network*, mechanical, chemical, corrosion, physical, electrical, and electrochemical properties of materials; *Navy News and Undersea Technology*, U.S. Navy contracts; *NNDC Addresses*, nuclear physics researchers; *NNDC Newsletter*, news and information of interest to the nuclear physics research community; *Nuclear Data (NUDAT)*; *PASCAL*, alternative name: *Programme Applique a la Selection et a la Compilation Automatiques de la Literature (PASCAL)*, science and technology; *Physical Property Data Service (PPDS)*; *Physics Abstracts (PA)*; *Physics Briefs (PB)*; *Physics Codes (PHYSCO)*, physics codes for the calculation of physics quantities; *Pollution Abstracts*; *PRODIS*, the humanization of labor; *PsycFILE*, industrial and organizational psychology; *Publications of the Institute for Research in Construction (IRCPUBS)*; *Raumordnung, Städtebau, Wohnungswesen, Bauwesen (RSWB)*, regional planning, town planning, housing, building construction, and civil engineering; *Recent Advances in Manufacturing (RAM)*; *REGLER*, alternative name: *REGISTER*, Swedish regulations and standards covering building, energy saving, environmental technology, physical planning, and urban technology; *RESAGRI*, agriculture; *Reuter TEXTLINE*, company and industry coverage; *ROADLINE*, former name: *GEOROAD*; roads, traffic, vehicles, road users, safety, and related topics; *SAE Global Mobility Database*, former name: Society of Automotive Engineers Abstracts; *SAE Abstracts*; *SCISCAN*, natural, physical, earth, environmental, biomedical, and life sciences; *SciSearch*, natural, physical, earth, environmental, biomedical, and life sciences; *Searchable Physics Information Notices (SPIN)*; *SEBAN Data Base*, scientific research; *Solid State and Superconductivity Abstracts*; *Standards and Directories/ Normes et Répertoires*, standards, directories, and publications covering occupational health and safety, electrical, mechanical, and mine safety, and environmental sciences; *Sugar Industry Abstracts*; *SVERKER*, professional senior engineers in Sweden offering expert advice on building and civil engineering; *System for Documentation and Information in Metallurgy (SDIM)*; *TECHNO-SEARCH*, new products, research and development, and marketing trends in Japanese industry; *Tekniikan Aikakauskenti Indeks*i (*TALI*), engineering, forest products, mining and metallurgy, architecture, chemistry, physics, mathematics, geology, information processing, industrial economics, energy, and information science; *TITLE-SEARCH*, alternative name: *Technical Report Title Hotline*, general engineering; *Transportation Research Information Service (TRIS)*, former name: *TRIS-ON-LINE*, all aspects of transportation; *VDI-Nachrichten*, economic, scientific, and technical aspects of engineering; *Verfahrenstechnische Berichte (VtB)*, chemical engineering, process engineering, and related fields; *Wer Baut Maschinen*, alternative name: *Who Makes Machinery?*; *Wheat, Barley and Triticale Abstracts*; *Who's Who in Technology*; *World Patents Index (WPI)*; and *Zentralstelle Dokumentation Elektrotechnik Database (ZDE)*, alternative name: *Literaturdatenbank Elektrotechnik*, former name: Dokumentationsring Elektrotechnik (DRE).

Claims and Patents. An alphabetic listing of claims and patents databases (8) available as of 1992 follows: *APIPAT*, alternative name: *Index to API Abstracts/ Patents*, former name: Index to Abstracts of Refining Patents, petroleum and energy industries worldwide; *BNA Daily News*, U.S. governmental and regulatory developments, policymaking, legislative activities, and legal issues; *BNA Patent, Trademark, and Copyright Daily*, U.S. patent, copyright, trademark, and unfair competition law developments; *BNA's Patent, Trademark, and Copyright Journal*, patent, trademark, and copyright law; *BREV*, all sectors of patentable activities in Belgium; *British Library Catalogue: Science Reference and Information Service*, all areas of science and technology, business and industrial property, including patents, trademarks, and registered designs; *Canadian Patent Index*; *Canadian Patent Reporter (CPR)*, Canadian patent and copyright legal decisions; *Chinese Patent Abstracts in English Data Base*; *CIB*, international patent classifications; *CIBEPAT*, patents and utility models registered in Spain; *CLAIMS Compound Registry*, former name: Class Code, Assignee, Index, Method Search—Compound Registry, chemical compounds; *CLAIMS/CITATION*, U.S. and some foreign patents in all subject areas; *CLAIMS/CLASS*, former name: Class Code, Assignee, Index, Method Search—Classification; *CLAIMS Reference*, U.S. patent classification and authority data; *CLAIMS/ Comprehensive Data Base*, U.S. patents in the fields of chemistry and engineering; *CLAIMS/ Reassignment & Reexamination*, patent reassignments and reexaminations; *CLAIMS/ U.S. Patent Abstracts*, alternative name: *CLAIMS Biblio/ Abstracts*, former name: Class Code, Assignee, Index, Method Search—General, Electrical, Mechanical and Chemical; *CLAIMS/ UNITERM*, U.S. patents in the fields of chemistry and engineering; *CLINPAT*, patent classifications; *Current Patents Evaluation*, patents covering cardiovascular, central nervous system, and antimicrobial therapies; *Current Patents Fast-Alert*, pharmaceutical patents; *Database on Legal Precedents Regarding Intellectual Property Rights*; *Deutsche Patent Datenbank (PATDPA)*, German patents and utility models; *ECLATX*, former name: *INPI-4*, the International Patent Classification scheme; *EDOC*, former name: *INPI-3*, European patents; *EPAT*, alternative name: *European Patent Register*, former name: *INPI-2*, European patents; *FPAT*, former name: *INPI-1*; *INPI BREVETS*, French patents; *IMSWorld Online Service*, the international pharmaceutical industry; *INPADOC Data Base (IDB)*, alternative name: *Patent Family Service/ Patent Register Service (PFS/ PRS)*, patents issued worldwide, including patent families and legal status; *International Economic Law Documents*; *ITALPAT*, Italian patent and model applications; *Japio*, Japanese patent, utility model, design, trademark, etc; *JURINPI*, French legal decisions related to patents and trademarks; *KPTN*, alternative name: *Korean Patents*; *KUMO*, alternative name: *Korean Utility Model*, Korean utility model specifications; *LEXIS Federal Patent, Trademark, and Copyright Library*, U.S. patent, trademark, and copyright law; *LEXPAT*, U.S. patents in all subject areas; *LitAlert*, U.S. patent and trademark litigation; *ORBPAT*, patents databases available online through the ORBIT Search Service; *PAPERCHEM*, scientific and technical aspects of the pulp and paper industries; *PATDATA*, U.S. patents in all fields of medicine, science, and engineering technology; *Patent Status File*, U.S. patent status changes; *PATGRAPH*, chemical formulas derived from patents; *PATOS European Legal Status and Alterations*, European patents and patent applications; *PATOS European Patent Applications*; *PATOS European Patents*; *PATOS German Patent Applications and Utility Models*; *PATOS German Patents*; *PATOS PCT Applications*, international patent applications; *PD-basen*, alternative name: *Patentdirektorsbasen*, Danish patents, trademarks, and designs, including pending applications; *PHARMSEARCH*, pharmaceutical patent information published in France, the United States, and by the European Patent Office (EPO); *SITADEx*, patents, trademarks, and designs registered in Spain; *SITADIN*, new law patents and utility models registered in Spain, including European patents designating Spain; *TRANSIN*, technology transfer, patents, new products and inventions, industrial know-how; *U.S. Patents*; *USCLASS*, alternative name: *U.S. Classifications*, U.S. patents in any subject area; *WESTLAW Intellectual Property Library*, former name: *WESTLAW Copyright, Patent and Trademark Library*, U.S. intellectual property law, covering copyrights, patents, and trademarks; *WESTLAW Texts and Periodicals Library*; *World Patents Index (WPI)*; and *WPI/ APIPAT*, petroleum and energy industries worldwide.

BIBLIOGRAPHY

1.

K. Y. Marcaccio, K. Hillstrom, C. Tomassini, and G. E. Turecki, eds., and M. E. Williams, founding ed., *Computer Readable Databases, A Directory and Data Sourcebook*, 8th ed., Gale Research Inc., Detroit, Mich., 1992; *Online Services: 1992 Review, Trends & Forecast*, SIMBA Information Inc., Wilton, Conn., 1992.

2.

A. Y. Chamis, *Vocabulary Control & Search Strategies in Online Searching*, *New Directions in Information Management Service*, No. 27, Greenwood Press, Westport, Conn., 1991; E. Auster, ed., *The Online Searcher*, Neal-Schuman, New York, 1990; C. H. Fenichel and T. H. Hogan, *Online Searching: A Primer*, Learned Info., Oxford, UK, 1990; P. J. Vigil, *Online Retrieval: Principles & Systems*, *Information Sciences Service*, John Wiley & Sons, Inc., New York, 1988; J. Schiele and I. Kurtev, *INIS Reference Service*, UNIPUB, 1984. See also M. E. Williams and F. Gibbs, eds., *Online Review, The International Journal of Online Information Systems*. *Online review* is the specialist journal for the online information industry; D. I. Raitt, ed., *The Electronic Library*. *The Electronic Library* is an international journal that focuses on the impact of computerized storage, cataloging, and retrieval applications in libraries and in information centers; *Monitor: An Analytical Review of Current Events in the Online and Electronic Publishing Industry*. *Monitor*, a monthly

Kirk-Othmer Encyclopedia of Chemical Technology (4th Edition)

460

newsletter, provides informed descriptions and analyses of events in the sphere of information transfer via electronic media; *Going Online: An Introduction to the World of Online Information*, Learned Information, Ltd., Oxford, UK. An instructional video, *Going Online* presents the concepts of online searching of the computer for the online novice.

3. Technical data, Information Market Indicators, Inc., Monticello, Ill., 1992.
4. R. T. Martinott, *Comp. Chem.* **1**(2), 12–18 (June 1992); R. T. Martinott, *Comp. Chem.* **1**(1), SR3–SR15 (April 1992); *Database Catalog*, DIALOG Information Services, Inc., Palo Alto, Calif., 1992, pp. 3–4; S. Lewis, *Plugging In—The Microcomputerist's Guide to Telecommunications*, Chilton's Computing Series, Chilton Book Co., Radnor, UK, 1984.
5. J. N. Daigle, *Queuing Theory for Telecommunications*, Addison-Wesley Series in Telecommunications, Addison-Wesley, New York, 1992; E. A. Edis and J. E. Varall, *Newnes Telecommunications Pocket Book*, Butterworth-Heinemann, UK, 1992; J. Van Duuren, P. Kastelein, and F. Schoute, *Telecommunication, Networks & Services*, Addison-Wesley, New York, 1992; *Telecommunications in the Age of Information*, Gordon Press, New York, 1992; S. Taylor, *Telecommunications on the MAC*, MIS PRESS, New York, 1992; J. Nellist, *Understanding Telecommunications & Lightwave Systems*, Inst. Electrical, Piscataway, N.J., 1992; M. Carpentier, S. Farnoux-Toporkoff, and C. Garric, *Telecommunication in Transition*, John Wiley & Sons, Inc., New York, 1992; W. Ritchie and V. Stern, eds., *Telecommunications Local Networks*, Van Nostrand Rheinhold, New York, 1992; R. K. Heldman, *Global Telecommunications: Layered Networks' Layered Services*, McGraw-Hill, San Francisco, 1992; J. C. Dvorak and N. Anis, *Dvorak's Guide to PC Telecommunications*, Osborne-McGraw, New York, 1992; K. Froelich, *Encyclopedia of Telecommunications*, Marcel Dekker, New York, 1992; L. S. Gross, *Telecommunications: An Introduction to Electronic Media*, Wm C. Brown Comm., Dubuque, Iowa, 1992. See also, N. Anis, *Complete Q Modem Toolkit: Everything You Need to Go On-Line*, Brady Computer Books, Prentice Hall, New York, 1992; L. Free, *PC Magazine Guide to Modem Communications*, Ziff-Davis, Dallas, Tex., 1992; J. Pournelle and M. Banks, *Pournelle's PC Communications Bible: The Ultimate Guide to Productivity with a Modem*, Microsoft, Redmond, Wash., 1992.
6. R. D. Athey, Jr., *Met. Finish.* **89** (2), 76 (1991).
7. M. Levene, *The Nested Universal Relation Database Model, Lecture Notes in Computer Science Service*, Vol. 595, Springer-Verlag, Heidelberg, 1992; "Database Programming Languages: Bulk Types & Persistent Data," *The Third International Workshop*, Aug. 27–30, Nafplion, Greece, 1991; J. M. Davis, *The Essential Guide to Database Marketing*, McGraw-Hill, New York, 1992; R. Hogan, *Managing IMS Databases*, QED Information Science, Wellesley, Mass., 1992; A. K. Elmagarmid, ed., *Database Transaction Models for Advanced Applications*, Series in Data Management Systems, Morgan-Kaufmann, Menlo Park, Calif., 1992; S. Khoshafian, *Object Oriented Databases*, John Wiley & Sons, Inc., New York, 1992; B. D. Klierer, *Database Modeling in a PC Environment*, Bantam Books, Inc., New York, 1992; K.-J. Raiha, *Design of Relational Databases*, Addison-Wesley, New York, 1992; P. Loucopoulos and R. Zicari, *Conceptual Modeling, Databases, & Case: An Integrated View of Information Systems Development*, John Wiley & Sons, Inc., New York, 1992; D. J. Harper, ed., "Specifications of Database Systems," *First International Workshop on Specifications of Database Systems, Glasgow, July 3–5, 1991*, Springer-Verlag, Heidelberg, 1992; K. Parsaye and M. Chignell, *Intelligent Database Tools & Applications: Object-Oriented Hypermedia, Visualization & Automatic Discovery*, John Wiley & Sons, Inc., New York, 1992.
8. File 230, DIALOG, Gale Research, Inc., Detroit, Mich., 1992.

General References

DIALOG Focus on Biotechnology, DIALOG Information Services, Inc., Palo Alto, Calif., Sept. 1989.

CASSI Cumulative (Chemical Abstracts Service Source Index), Chemical Abstracts Service, American Chemical Society, Columbus, Ohio, 1991.

M. E. Williams, "The State of Databases Today: 1993," in K. Y. Marcaccio, K. L. Nordan, and G. E. Turecki, eds., *Gale Directory of Databases*, Gale Research, Inc., Detroit, Mich., 1993.

K. E. Marcaccio, K. Hillstrom, C. Tomassini, G. E. Turecki, and M. E. Williams, eds., *Computer Readable Databases: A Directory and Data Sourcebook*, 8th ed., Gale Research, Inc., Detroit, Mich., 1992.

M. E. Williams, *Information Market Indicators: Information Center/Library Market*, issue 39, Information Market Indicators, Monticello, Ill., 1992.

Martha E. Williams
University of Illinois

James L. Grant
DIALOG Information Services, Inc.

DATA INTERPRETATION AND CORRELATION.

See CHEMOMETRICS; ENGINEERING, CHEMICAL DATA CORRELATION.

DDT.

See INSECT CONTROL TECHNOLOGY.

DECAHYDRONAPHTHALENE, DECALIN.

See NAPHTHALENE.

DECARBOXYLASES.

See ENZYME APPLICATIONS, INDUSTRIAL; MICROBIAL TRANSFORMATIONS.

DEFLOCCULATING AGENTS.

See CCLAYS; DISPERSANTS; FLOCCULATING AGENTS.

DEFOAMERS

The control or elimination of the foam that occurs in many industrial processes is often a key factor in their efficient operation. The additives that are used in low concentration to achieve this effect are known variously as defoamers, antifoaming agents, foam inhibitors, and foam controllers. Defoaming implies breaking a pre-existing foam whereas antifoaming or foam inhibition indicates prevention of the formation of that foam. Such distinctions call for different product features. A defoamer is expected to exhibit rapid knockdown of a foam, whereas longevity of action might be the key requirement in many antifoam applications. Despite these varying performance features, many applications require both preventive and control functions, and in practice the same types of materials are used both for antifoaming and defoaming. For this reason, the general term defoamers, as used in this article, is meant to encompass all product types and degrees of action encountered with such process aids. The topic is sometimes known as chemical (physicochemical would be better) foam control as opposed to thermal and mechanical foam control, such as centrifuging or spraying of the foam or the use of ultrasound, which are not covered here.

Many industries now rely on the efficient and economical use of defoamers both as a process aid in product manufacture and to increase the quality of the finished product in its subsequent application. The most obvious use of defoamers as process aids is to increase holding capacity of vessels and improve efficiency of distillation or evaporation equipment. They are also used to improve filtration, dewatering, washing, and drainage of suspensions, mixtures, or slurries. Examples of industrial operations that benefit in these ways from the use of defoamers include oil well pumping; gas scrubbing at petrochemical plants; polymer and chemical synthesis and processing, particularly in monomer stripping; textile dyeing and finishing; leather processing; paint and adhesive manufacture; phosphoric acid production; control of wastewater and sewage; food preparation, notably the refining of sugar; the brewing of beer; and penicillin production by fermentation. Among the finished products that are improved in quality or efficacy by the proper inclusion of defoamers are lubricants, particularly cooling lubricants in metal working; diesel fuel, hydraulic and heat-transfer fluids; paints and other coatings; adhesives; inks; detergents; and antifatulence tablets.

The use of vegetable and mineral oils as defoamers has been known for a long time. However, most modern defoamers are complex, formulated specialty chemicals. They are usually proprietary products and the patent literature is the best guide to their probable composition. Patent surveys (1,2) are particularly valuable sources of information. Other useful reviews include those in encyclopedias and handbooks (3–5), some of which have a special focus on defoamers in coatings (4) and on polymer aspects of defoaming (5). Companies involved in the production of defoamers range from large, basic polymer producers to small regional formulators. In addition to control of foam and associated features such as rate of foam knockdown and the persistence of the effects, other frequently needed application requirements of these specialty materials include adequate shelf life, absence of adverse effects on and by the products being treated, ease of handling, lack of toxicity to manufacturing personnel and users, environmental acceptability, and cost-effectiveness. Defoamers range from relatively inexpensive mineral oils to costly fluorinated polymers but it is not the cost per kilogram of defoamer that matters, but rather the cost per unit produced using this processing aid. Another factor that strongly influences the choice of a specific defoamer is its ancillary surface properties such as wetting, dispersion, and leveling.

Classification of Defoamers

Modern defoamers typically contain numerous ingredients to meet the diverse product requirements for which they are formulated. A variety of classification schemes is possible including classification by application, physical form, and chemical type. Classification by chemical type is a very satisfactory approach. In general defoamers contain a variety of active ingredients in both the solid and liquid states and numerous ancillary agents such as emulsifiers, spreading agents, thickeners, preservatives, carrier oils, compatibilizers, solvents, and water. Not all defoamers contain all classes of components; some complex formulations contain several compounds in some of the categories.

ACTIVE INGREDIENTS

These are the components of the formulation that do all or most of the actual foam control work. Traditionally, defoamers were single component liquids or homogeneous solutions of vegetable or mineral oils, but more recently a number of active hydrophobic solids have been utilized so effectively that in a dispersion of hydrophobic solids in a traditional oil such as castor oil [8001-79-4], the oil could be classed as a carrier oil rather than an active ingredient.

Liquid-Phase Components. It is usual to classify organic liquids by the nature of the polar or hydrophilic functional group, ie, alcohol, acid, ester, phosphate, etc. Because lowering of surface tension is a key defoamer property and since this effect is a function of the nonpolar portion of the liquid-phase component, it is preferable to classify by the hydrophobic, nonpolar portion. This approach identifies four liquid phase component classes: hydrocarbons, polyethers, silicones, and fluorocarbons.

Many similar hydrocarbon fluids such as kerosene and other paraffinic and naphthenic mineral oils and vegetable oils such as linseed oil [8001-26-1], corn oil, soybean oil [8001-22-7], peanut oil, tall oil [8000-26-4], and castor oil are used as defoamers. Liquid fatty alcohols, acids and esters from other sources and poly(alkylene oxide) derivatives of oils such as ethoxylated rosin oil [68140-17-0] are also used. Organic phosphates (6), such as tributyl phosphate, are valuable defoamers and have particular utility in latex paint applications. Another important class of hydrocarbon-based defoamer is the acetylenic glycols (7), such as 2,4,7,9-tetramethyl-5-decyne-4,7-diol which are widely used in water-based coatings, agricultural chemicals, and other areas where excellent wetting is needed.

Numerous organic polymers have also been proposed as foam control agents (5); these include polyisobutylene [9003-27-4], poly(alkyl acrylates), polyalkylene polyamines, polyalkyleneimines, and a variety of other copolymers and low molecular weight resins. Poly(alkylene oxide) homopolymers and copolymers are frequently encountered in liquid-phase antifoam components. For example, copolymers [106392-12-5] of poly(ethylene oxide) and poly(propylene oxide) are used to reduce foam in the acid–gas scrubbing process (8). High molecular weight adducts of propylene oxide and polyhydric alcohols such as glycerol [56-81-5] and pentaerythritol [115-77-5] have also been reported to have useful antifoaming properties (3). Sometimes the polyether is mixed with other liquids, eg poly(propylene oxide) [25322-69-4] and polydimethylsiloxane [9016-00-6]; sometimes the two materials are copolymerized.

Silicone oils are particularly effective antifoaming agents in nonaqueous systems because of their low surface tension and incompatibility. Polydimethylsiloxane (PDMS) is the most important silicone and is widely used in products for the petroleum industry, where its good thermal stability is very advantageous (9). Other useful silicone defoamers are polytrifluoropropylmethylsiloxanes [25791-89-3] and other fluorosilicones, which are very effective in nonaqueous foaming systems such as organic solvents and crude hydrocarbon stocks. Fluorocarbons are the most expensive class of antifoam fluid. Fluorocarbon oils and fluorine-containing amides such as the *N*-(alkylamino-trimethylene)perfluoro-octanamides are used as antifoaming additives to lubricants and jet fuels (10).

Solid-Phase Components. Dispersed solids are vital ingredients in commercial antifoam formulations. Much of the current theory on antifoaming mechanism ascribes the active defoaming action to this dispersed solid phase with the liquid phase primarily a carrier fluid, active only in the sense that it must be surface-active in order to carry the solid particles into the foam films and cause destabilization. For example, polydimethylsiloxane, despite its considerable effectiveness in nonaqueous systems, shows little foam-inhibiting activity in aqueous surfactant solutions. It is only when compounded with hydrophobic silica [7631-86-9] to give the so-called silicone antifoam compounds that highly effective aqueous defoamers result. Using the same classification adopted for liquids, gives three solid phase component classes: hydrocarbons, silicones, and fluorocarbons.

A variety of waxy hydrophobic hydrocarbon-based solid phases are used including fatty acid amides and sulfonamides, hydrocarbon waxes such as montan wax [8002-53-7], and solid fatty acids and esters. The amides are particularly important commercially. One example is the use of ethylenediamine distearamide [110-30-5] as a component of latex paint and paper pulp black liquor defoamer (11). Hydrocarbon-based polymers are also used as the solid components of antifoaming compositions (5); examples include polyethylene [9002-88-4], poly(vinyl chloride) [9002-86-2], and polymeric ion-exchange resins.

In most cases, these active defoaming components are insoluble in the defoamer formulation as well as in the foaming media, but there are cases which function by the inverted cloud-point mechanism (3). These products are soluble at low temperature and precipitate when the temperature is raised. When precipitated, these defoamer–surfactants function as defoamers; when dissolved, they may act as foam stabilizers. Examples of this type are the block polymers of poly(ethylene oxide) and poly(propylene oxide) and other low HLB (hydrophilic–lipophilic balance) nonionic surfactants.

Pure silicone solids, such as silicone resins, are used in defoamer formulations, but the key material in this category is hydrophobic silica, listed here because the hydrophobic treatment is usually a silicone polymer or silane monomer. The development of these hydrophobic silicas (12) is widely acknowledged as one of the most significant advances in antifoam technology. They are used with a considerable variety of liquid components including hydrocarbons, polyethers, and silicones. The three most common ways of preparing the hydrophobic silica are to spray the silica with silicone oil and heat at 250–350°C, to treat with organochlorosilane vapors in an autoclave, and to disperse the silica in a silicone oil at elevated temperatures. Hydrophobic silica can also be produced by treatment with alcohols, fatty amines and hydrocarbon waxes. The application of hydrophobic silica formulations as antifoaming agents has been reviewed with an emphasis on mechanism and use in textile dyeing (13).

Fluorocarbon solids are rare in defoamer compositions, presumably on account of their cost. Solid fluorine-containing fatty alcohols and amides are known. The most familiar fluorocarbon solid is polytetrafluoroethylene [9002-84-0]. Because it is more hydrophobic than silicone-treated silica, it might be expected to perform impressively as a defoamer component (14). However, in conventional hydrocarbon oil formulations it works poorly because the particles aggregate strongly together. In lower surface tension fluids such as silicone and fluorocarbon oils, the powdered polytetrafluoroethylene particles are much better dispersed and the formulation performs well as a defoamer.

ANCILLARY AGENTS

Surface-Active Materials. Since defoaming is a surface effect, all active components are necessarily surface active materials in that sense, but in this ancillary category the surfactants are incorporated in the formulation for other effects such as emulsification or to enhance dispersion. Emulsifiers are essential in the common oil-in-water emulsion systems but they are also required where mixtures of active liquid components are used. For example, specialized oil-in-oil emulsifiers are needed in defoamers based on silicone polyether mixtures. Oil-in-water emulsifiers are incorporated in some defoamers even when the final product contains no water. This is to promote emulsification (self-emulsifiable) or dispersion into aqueous foaming systems. These additives increase the speed of foam decay by promoting rapid dispersion of the defoamer throughout the foaming media. Examples of emulsifying agents used in defoamer compositions are fatty acid esters and metallic soaps of fatty acids; fatty alcohols and sulfonates, sulfates and sulfosuccinates; sorbitan esters; and ethoxylated products such as ethoxylated octyl or nonylphenols.

Carriers. The function of the carrier is to provide an easily handleable, readily dispersible system for delivering the active defoamer components to the foaming system and also to tie the complex defoamer formulation together, ie, coupling agents, compatibilizers, or solubilizers. Sometimes the carrier is used simply as an extender to lower the cost of the final product. Many of the low viscosity organic solvents that are used also exhibit some antifoaming properties in cases where they are both insoluble and of lower surface tension than the medium to which they are applied. Any of the usual paraffinic, naphthenic, aromatic, chlorinated or oxygenated organic solvents can be used, but aliphatic hydrocarbons are the most common. Water is often used as a carrier fluid. In these cases, the defoamer product is typically an oil-in-water emulsion. With growing concern over unrecovered solvents, this has become a preferred type of defoamer formulation. Such products usually require preservatives to prevent bacterial spoilage in the drum or other shipping and storage container.

Sometimes the defoamer is required in a solid form; for example, to be suitable for incorporation into a low-sudsing detergent powder or agricultural chemical composition. Water soluble inorganic sorbent carriers such as sodium sulfate [7757-82-6], sodium carbonate [497-19-8] or sodium tripolyphosphate [7758-29-4] are used as well as organic polymers such as methylcellulose [9004-67-5]. Sometimes the particles are further encapsulated with a coating that preserves the integrity of the defoamer formulation in storage with the detergents but allows disruption on contact with water in the wash process. When both the antifoam and the solid support are food-grade materials, they can also be used in food processing, brewing, and pharmaceutical areas. Pre-measured water-dissolvable packets of this type of product offer added convenience.

Commercial Sources

The defoamer market is large and very specialized; suppliers differ markedly in orientation and in range of product lines. Some suppliers are large, international, polymer producers with a basic manufacturing capability for the specific type of defoamer material they supply, whereas others are small, often regional, formulators who focus on the concerns of particular industries. The diversity of product type and application has resulted in hundreds of such suppliers, some with very extensive product lines. Consequently, it is impossible to present a comprehensive listing. Table 1 is a short selection illustrating this diversity. It is primarily based on tables given in References 3 and 5, and reflects commercial product availability in the United States and Europe. Trade name selections have been limited to one per supplier.

Table 1. Commercial Defoamer Examples

Trade name	Supplier	Use or composition
Airex	Th. Goldschmidt AG	silicone–organic defoamers for paints and coatings
Antimussol	Sandoz AG	textile and paper industries, cutting oils
Antispumin	Chemische Fabrik Stockhausen GmbH	paper, textile, and sugar industries
Baysilone-Entschaumer	Bayer AG	silicone defoamers
Bubble Breaker	Witco Corp.	inks and coatings foam control agents
Defoamer Crystals	Imperial Manufacturing Co. Inc.	solid antifoams
Dow Corning Antifoam	Dow Corning Corp.	silicone defoamers
Foamaster	Henkel KGaA	coatings and other industries
Fumexol	Ciba-Geigy AG	textile industry
Hallcomid	The C.P. Hall Co.	dimethyl fatty acid amides
Mazu	PPG-Mazer Chemicals, Inc.	wide variety of silicone and ethoxylated surfactant types
Nalco	Nalco Chemical Co.	coatings industry
Pegosperse	Glyco Chemicals Inc.	ethoxylated ester oils
Pluronic	BASF AG	poly(ethylene oxide- <i>co</i> -propylene oxide)
Polyglycol	Dow Chemical Co.	poly(propylene oxide) used in latex and emulsion paints
Ralulac	Raschig AG	emulsion paints, leather and textile industries
Sentry Simethicone	Union Carbide Corp.	silicone defoamers for antifatulent preparations
Silcolapse	Imperial Chemical Industries PLC	silicone defoamers
Surfynol	Air Products and Chemicals Inc.	acetylenic glycols: defoamers and low foam wetting agents

Defoaming Theory

Foams are thermodynamically unstable. To understand how defoamers operate, the various mechanisms that enable foams to persist must first be examined. There are four main explanations for foam stability: (1) surface elasticity; (2) viscous drainage retardation effects; (3) reduced gas diffusion between bubbles; and (4) other thin-film stabilization effects from the interaction of the opposite surfaces of the films.

The stability of a single foam film can be explained by the Gibbs elasticity *E* which results from the reduction in equilibrium surface concentration of adsorbed surfactant molecules when the film is extended (15). This produces an increase in equilibrium surface tension that acts as a restoring force. The Gibbs elasticity is given by equation 1 where σ is surface tension and *A* is surface area of the film.

$$E = 2A d\sigma / dA$$

(1)

In a foam where the films are interconnected the related time-dependent Marangoni effect is more relevant. A similar restoring force to expansion results because of transient decreases in surface concentration (increases in surface tension) caused by the finite rate of surfactant adsorption at the surface.

Such nonequilibrium surface tension effects are best described in terms of dilatational moduli thanks to developments in the theory and measurement of surface dilatational behavior. The complex dilatational modulus ϵ^* of a single surface is defined in the same way as the Gibbs elasticity as in equation 2 (the factor 2 is halved as only one surface is considered).

$$\epsilon^* = A d\sigma / dA$$

(2)

In a dilatational experiment, where the surface is periodically expanded and contracted, ϵ^* is a function of the angular frequency (ω) of the dilatation as in equation 3 where ϵ_d is the dilatational elasticity and η_d is the dilatational viscosity.

$$\epsilon^*(i\omega) = \epsilon \cos\theta + i\epsilon \sin\theta$$
$$= \epsilon_d(\omega) + i\omega\eta_d(\omega)$$

(3)

It has been shown (16) that a stable foam possesses both a high surface dilatational viscosity and elasticity. In principle, defoamers should reduce these properties. Ideally a spread duplex film, one thick enough to have two definite surfaces enclosing a bulk phase, should eliminate dilatational effects because the surface tension of an insoluble, one-component layer does not depend on its thickness. This effect has been verified (17). Silicone antifoams reduce both the surface dilatational elasticity and viscosity of crude oils as illustrated in Table 2 (17). The PDMS materials are Dow Corning Ltd. polydimethylsiloxane fluids, SK 3556 is a Th. Goldschmidt Ltd. silicone oil, and FC 740 is a 3M Co. Ltd. fluorocarbon profoaming surfactant.

Table 2. Dilatational Elasticities and Viscosities of Crude Oil at 1 mHz^a

Crude oil	Additive ^b identity	Additive viscosity, mm ² s(cST)	ϵ_d , mN / m (– dyn / cm)	η_d , mNs / m
North Sea	none		1.34	153
	PDMS	12,500	0.69	90
	PDMS	60,000	0.51	33
	SK 3556		0.99	87

Middle East	none	1.63	105
PDMS	60,000	1.19	53
	FC 740	4.36	377

^a Ref. 17.
^b Concentration = 1 ppm.

Both high bulk and surface shear viscosity delay film thinning and stretching deformations that precede bubble bursting. The development of ordered structures in the surface region can also have a stabilizing effect. Liquid crystalline phases in foam films enhance stability (18). In water-surfactant-fatty alcohol systems the alcohol components may serve as a foam stabilizer or a foam breaker depending on concentration (18). Alcohol : surfactant ratios less than that corresponding to the liquid crystalline phase enhance film stability; higher ratios produced by contact with alcohol droplets disrupt this phase and cause film instability. Liquid phase defoamer components may dilute or destroy such stabilizing phases or they may simply contribute lower surface shear viscosities than the foam stabilizing surfactant (profoamer). For example, the very low surface shear viscosity of polydimethylsiloxane (19,20) is often cited as a contributing factor to its effectiveness in defoamer compositions. On the other hand, it should be possible to form too rigid a surface that will be prone to rupture. Thus dilatational moduli that are too high also result in foam instability. This is the case in a study of diesel fuel antifoaming with a fluorosilicone antifoam agent (21). The marked viscoelasticity of this system is most unexpected.

Reduced gas diffusion between bubbles delays collapse by retarding bubble-size changes and the resulting mechanical stresses. Consequently single films persist longer than the corresponding foam, but it seems to be a minor factor in practical defoaming situations. The same is true of other thin-film stabilization effects from the interaction of the opposite surfaces of the film. These comprise both electric double-layer repulsion for ionic surfactants and entropic repulsion of polymer chains in the surface for nonionic materials. These effects are of paramount importance in determining the stability of very thin (less than 10 nm) films, but in practice the real challenge is usually to defoam films of at least several hundred nanometers in thickness where such effects have not begun to be significant.

All these mechanisms except high bulk viscosity require a stabilizer in the surface layers of foam films. Accordingly, most theories of antifoaming are based on the replacement or modification of these surface-active stabilizers. This requires defoamers to be yet more surface active; most antifoam oils have surface tensions in the 20 to 30 mN/m range whereas most organic surfactant solutions and other aqueous foaming media have surface tensions between 30 and 50 mN/m (— dyn/cm). This is illustrated in Table 3.

Table 3. Surface Tensions of Surfactants and Defoamers

Material	CAS Registry Number	Surface tension ^a mN/m (— dyn/cm)	Temperature, °C	References
<i>Surfactants</i>				
sodium lauryl sulfate C ₁₂ H ₂₅ SO ₄ Na ⁺	[151-21-3]	39.5	25	22
sodium 12-butoxydodecyl sulfate C ₄ H ₉ OC ₁₂ H ₂₄ SO ₄ Na ⁺	[3694-71-1]	44.0	25	22
lauryl pyridinium bromide C ₁₂ H ₂₅ C ₅ H ₅ N ⁺ Br	[104-73-4]	41.2	30	22
C ₁₂ H ₂₅ (OC ₂ H ₄) _n OH ^b	[9002-92-0]	36.3	23	22
<i>para</i> - <i>t</i> -C ₈ H ₁₇ C H ₄ (OC ₂ H ₄) _n OH ^c	[9036-19-5]	33.5	25	22
<i>Surfactant/ defoamer</i>				
Surfynol 104 ^d		31.4	25	7
Pluronic L62 ^e		42.8	25	23
<i>Defoamers</i>				
poly(oxypropylene), mol wt 3000		31.2	20	24
polydimethylsiloxane, mol wt 3900		20.2	20	24
kerosene	[8008-20-6]	27.5	29	25
mineral oil (MWP paraffin)	[8020-83-5]	28.8	20	26
corn oil	[8001-30-7]	33.4	20	27
peanut oil	[8002-03-7]	35.5	20	27
tributyl phosphate	[126-73-8]	25.1	20	28

^a Surfactant values are at the critical micelle concentration (CMC) in aqueous solution; surfactant + defoamer values are at 0.1% concentration in aqueous solution.
^b Laury end-capped polyoxyethylene.
^c Polyoxyethylene end-capped with substituted phenyl group.
^d Surfynol 104 is an acetylenic glycol, 2,4,7,9-tetramethyl-5-decyne-4, 7-diol [126-86-3] marketed by Air Products and Chemicals, Inc.
^e Pluronic L62 is a poly(oxyethylene)–poly(oxypropylene)–poly(oxyethylene) copolymer marketed by BASF AG.

Based on this low surface tension feature and the commonly observed insolubility of defoamers, two related antifoam mechanisms have been introduced (29): (1) The agent dispersed in the form of fine drops enters the liquid film between bubbles and spreads as a duplex film. The tensions created by this spreading lead to the rupture of the original liquid film. (2) A droplet of the agent enters the liquid film between bubbles, but rather than spreading produces a mixed monolayer on the surface. This monolayer, if of less coherence than the original film-stabilizing monolayer, causes destabilization of the film.

The spreading process is governed by the spreading coefficient *S* defined as in equation 4 (30) where ζ_f is the surface tension of the foaming medium, ζ_d the surface tension of the defoamer, and ζ_{df} the interfacial tension between them.

$$S = \zeta_f - \zeta_d - \zeta_{df}$$

(4)

The lower the value of ζ_d , the more likely it is that *S* is positive indicating a thermodynamic tendency for the process to occur. Longitudinal wave theory has been applied to the defoamer spreading process as in equation 5 where *P* is the penetration depth of a spreading droplet of initial radius *R*, viscosity η , and density ρ .

$$P = R(4\eta^2 - 3\rho D\Delta\sigma)^{-1/3}$$

(5)

D is the final layer thickness and $\Delta\zeta$ the change in surface tension during the passage of the wave. Spread insoluble films give low $\Delta\zeta$, ie, high penetration depth and maximum disruption. *P* can be of the order of ten times the droplet size.

The theories discussed so far adequately account for agents consisting only of insoluble liquids such as silicone fluids used in the defoaming of crude oil. However, practical experience also shows that dispersed hydrophobic solids can greatly enhance defoamer effectiveness in certain cases, particularly in aqueous foam systems. Materials such as hydrophobic silica (13) or high melting-point hydrocarbon amides such as ethylenediamine distearamide are notably effective. A strong correlation between defoamer action and contact angle for silicone-treated silica in a hydrocarbon oil has been demonstrated (32). Most recent publications on the mechanism of defoamers have concentrated on the role of the dispersed hydrophobic solids. There are a number of contradictory older hypotheses for their mode of action that have been discounted (33) and replaced with the idea that dewetting of the hydrophobic silica by the bubble film causes foam collapse by the direct mechanical shock of the event (33,34). Such dewetting helps thin the film and promote instability, and is particularly effective when sizes are such that the particle occupies both surfaces of the film. In this bridged situation, if the contact angle is large enough, a capillary pressure drop that pushes liquid away from the particle is developed (34). This flow results in migration of the air-liquid-solid interfaces towards each other until they meet and the film ruptures as it pinches off the particle. This concept offers an explanation for observed particle size and shape effects (14,34,35). Note also that liquid droplet antifoaming can be interpreted in terms of dewetting by the foam film rather than by the spreading mechanism (36), although liquid particles have been found to be much less effective than solid particles (35). A number of cinephotomicrographic studies are consistent with these dewetting ideas (37,38).

Other workers suggest similar arguments, but consider solvation of the silica by foam-stabilizing surfactant adsorption to be an important part of the mechanism (39). Not everyone agrees (13,40), and debate continues about whether it is the fluid, solid, or a synergistic combination that is the true locus of the antifoaming effects. These dewetting and solvation themes imply that the hydrophobic particles become detached from the carrier fluid. This separation accounts for the familiar experience of performance of antifoam compositions diminishing with time. Further additions of defoamer are needed to maintain effective foam control. Adsorption of antifoam components on other available surfaces, particularly in systems containing suspended solids, mutual saturation of foamer and defoamer changing the spreading and dewetting pressures, and coalescence of antifoam emulsions also contribute to this progressive loss of effectiveness.

Applications

The main industries and broad product groups that utilize defoamers are indicated by the bold face headings in this section. These examples illustrate the wide variety of defoamer applications.

Adhesives and Sealants. Most industrial adhesives contain surface active components and additives, and air entrainment during their mechanical application can significantly reduce joint strength. Defoamers are usually formulated into adhesives to protect users against such difficulties. Additional benefits, such as improved uniformity of products, increased throughput and reduced labor costs can also result from the use of defoamers during adhesive application. The footwear and nonwoven fabric industries are extensive users of defoamers in this way.

Chemical Processing. Agitation, distillation, and pressure differences are commonplace in many chemical processes. These are conducive to foam formation, and even when the plant design minimizes these problems it is still often necessary to employ defoamers. A review of this field (41) lists numerous problems encountered with unwanted foam in chemical processes: increased cost of coping with safety hazards from slippery floors, corrosive residues or flammable liquids; interference with process instruments, pumps, etc; slow drainage of liquids from products being dried; product rejection due to incompletely filled containers; reduced capacity in vats; premature failure in bearings and other mechanical devices due to loss of lubrication; separation and segregation of critical process ingredients; and perceived negative environmental impact and poor community relations on discharge.

Cleaning Compounds. A considerable growth area for defoamers is in the formulation of low-foaming detergents and cleaners. This is in response to the growing automation of cleaning equipment and the need to operate it most effectively and efficiently. Incorporation of defoamers in detergents in such a way that the low-foaming property does not drift with storage time has been a considerable challenge. The antifoam composition has to survive storage with the surfactants, builders, bleaches and other auxiliary agents, and yet function properly as soon as the detergent is added to the wash water. This application has been reviewed from a European perspective (42). The solution is to prevent the defoamer from migrating within the detergent powder matrix such that the droplet size of the defoamer released into the wash water is the same regardless of storage conditions. Numerous patent examples of the four general ways in which this can be achieved have been provided (42): encapsulation in a water-soluble or water-dispersible wax; adsorption onto a carrier prior to wax coating; microencapsulation with film-forming polymers; and adsorption onto specially prepared inorganic salts.

Construction. Polymer dispersions in cements, mortars, and plastics are being increasingly used in the construction industry (3). Their plasticizing effect allows reduced amounts of water to be used, and they also confer strength and adhesion benefits in certain situations. The emulsifiers and dispersants used in these products can cause air entrainment problems with a detrimental effect on the ultimate stability of the construction. Powdered defoamers that can be directly added to the cement are commonly used. They are usually high surface area, highly absorbent inorganic fillers that have been treated with liquid defoamer compositions, similar to the products used in low-foaming detergents. Another approach to this problem is for the latex manufacturers to formulate low-foaming polymer dispersions with appropriate defoamers having good long-term stability.

Fermentation Processes. The efficient production of penicillin, yeasts, and single-celled protein by fermentation requires defoamers to control gas evolution during the reaction. Animal fats such as lard [61789-99-9] were formerly used as a combined defoamer and nutrient, but now more effective proprietary products are usually employed. Defoamer application technology has also improved. For example, in modern yeast production facilities, the defoamers are introduced by means of automatic electrode-activated devices. One concern in the use of defoamers in fermentation processes is the potential fouling of membranes during downstream ultrafiltration (qv). Silicone antifoams (43,44) seem less troubled by this problem than other materials.

Fertilizers. The fertilizer industry for many years has used tall oil fatty acids for the production of phosphoric acid [7664-38-2] by the digestion of phosphate-containing rocks with sulfuric acid. Carbon dioxide is liberated and presents a difficult, highly acidic foaming problem. Formulated products, many of which continue to contain tall oil fatty acids but which also contain emulsifiers and wetting agents such as dioctyl sodium sulfosuccinate [577-11-7], which greatly increase their effectiveness, are now used. Partially wetted gypsum particles contribute to the foam stabilization, and it is the modification of the gypsum wettability by these wetting agents that makes them so effective in this application.

Food and Beverages. Defoamer applications in the food and beverage industry include uses in both the preparation and processing of foodstuffs and in the cleaning and disinfecting of containers. The sugar beet industry is a prolific user of defoamers. This is now a fully mechanized automated procedure. The sugar is extracted with hot water, treated with limewater and carbon dioxide, filtered, and the filtrate subjected to evaporation. Foaming occurs in many of these steps. Another food processing area with considerable foam problems involves potatoes. Wash baths are universal in the production of chips, fries, mashed potatoes and potato starch. Proteins and other natural products in potatoes cause these troublesome foams. Starch foam in particular provides a considerable defoaming challenge. Beers and wines are also produced with the aid of defoamers. They permit more efficient use of such things as vats and containers, and permit more controllable bottling procedures. The growth in returnable bottles and the widespread use of automated mechanical cleaning equipment has increased use of defoamers in this industry. They can be used directly or incorporated in low-foaming cleaning agents.

Leather. Almost every stage in leather processing from the initial rawhide preparation, through tanning and dyeing and other finishing treatments, has the potential for causing foaming difficulties. Many of the preparations used in these steps contain wetting agents and other surfactants, or are applied in the form of emulsions or dispersions. The trend, as with many other defoamer applications, is to formulate the defoamer into the treatment product rather than deal with numerous optimized defoamers at each of the several processing steps.

Metal Working. The metal working industry encounters considerable foaming problems with the cutting oils and coolants that are sprayed onto the tool-workpiece interface to provide cooling, controlled lubrication, corrosion protection and increased tool life. These coolants are provided as mineral oil emulsifier concentrates that are diluted with water by the user. For many years this industry tolerated the foam difficulty in its many milling, drilling, grinding, rolling, and drawing operations but now uses defoamers such as formulated silicones and dispersions of fatty amides in mineral oils. Calcium soaps are also used as foam inhibitors in coolants. They are formed as a finely divided suspension of insoluble particles when the soaps present in the cooling lubricant concentrate react with hardening agents added to the dilution water.

Oil and Petrochemicals. There are a variety of uses for defoamers in oil recovery. They are used in some of the materials used in oil extraction, such as in drilling muds and cement lining, and also directly with the crude oil itself. In its natural state crude oil contains dissolved gases held by high reservoir pressure. When this live crude oil is extracted and passed into the low pressure environment of a gas-oil separator, the dissolved gases are

liberated and can cause troublesome foam that leads to oil losses via the gas stream and downstream equipment damage (45). Foaming is a problem in other petrochemical operations including distillation, cracking, coking and asphalt processing. A specific example (44) is the extraction of benzene [71-43-2], toluene [108-88-3], and xylenes [1330-20-7] (see BTX PROCESSING) from aromatic-rich feedstock using tetramethylenesulfone [126-33-0] (Sulfolane). One of the more recent advances in the downstream petroleum market has been the introduction of defoamers in diesel fuel (46).

Coatings. Foam problems occur both during the preparation of paints and coatings and in their application. The use of ball mills and other equipment for pigment dispersion provides ideal conditions for mixing in air, and the presence of surfactants in the formulations assures considerable persistence of the foams that are generated. These problems may be controlled during manufacture by mechanical means, but a defoamer is almost always required for foam control during application since application methods vary considerably, for example, roller, brush, dip, or spray methods. The proper choice and minimum use of surfactants such as dispersants, flow agents, and wetting agents can minimize but not eliminate the use of defoamers to prevent surface defects during application such as cratering, pinholes, fisheyes and orange peel. The use of defoamers in coatings has been surveyed (4) and pictures of these various defect types presented. In addition to the final dried film appearance, the defoamer must not detract from other properties such as color acceptance, gloss, and adhesion.

Polymers. Foam is often a particular problem in the production of polymers (5). There are numerous situations where foam can reduce the production capacities of vats and vessels and cause problems in pumps, meters, and other equipment, particularly distillation and evaporation equipment. Foam is frequently a problem when stripping off a monomer from a polymer. Examples are in the production of styrene–butadiene [9003-55-8] and acrylonitrile–butadiene [9003-18-3] rubber latices. These latices are stabilized by surfactants that greatly contribute to foaming difficulties. Another problem foam area is in the stripping of unreacted monomer from poly(vinyl chloride) suspensions. In this process, vinyl chloride [75-01-4], a gas at room temperature, is liquefied by pressure, emulsified in water with surfactants and catalysts, and heated to bring about polymerization. The recovery of unpolymerized monomer by distillation from this mixture produces a severe foaming problem.

Pulp and Paper. The pulp and paper industry presents some of the most critical and troublesome foam problems, and this industry has led the way in the use of defoamers. Early use of large amounts of kerosene or fuel oil has given way for ecological and cleanliness reasons to much more effective formulated defoamers. Foams are encountered at every stage from pulping, through paper fabrication and coating, to printing. A variety of wastewater streams are generated that are very prone to foaming because of the presence of dissolved soaps. Specific defoamer products are often tailored for each different stream. The so-called black-liquor defoamers were the first hydrophobic silica in hydrocarbon oil products that were then extended to other industries. The use of these defoamers has allowed some Kraft pulp mills to exceed original designed capacity.

Textiles. Defoamers are required in the jet dyeing of textiles (13,44,47). This process, which is mainly used with polyester fibers, is carried out at elevated temperature and pressure, and involves pumped recirculation of the dyeing medium from the reservoir through the jets. When the operation is complete and the pressure released, severe foaming can result in the absence of an effective defoamer. Foam control is needed in other dyeing processes also (13), such as continuous dyeing and beck dyeing. For example, in the dyeing of knitted fabrics of textured polyester fibers, foaming of the dye liquor can cause the material to float resulting in uneven application of the dye. Various surfactant mixtures are used in this area and bring additional benefits such as wetting and solubilization as well as foam control. Acetylenic glycols are very useful in this context (7). Dyeing is only one of several steps where foaming can cause difficulties with fiber processing. Defoamer may be required when any size or finish is applied to a textile material. One example is the pretreatment step in the processing of cotton. The fabric is exposed to strongly alkaline sodium hydroxide and anionic surfactant solution. Phosphate ester defoamers are much used in this application. Textile operations are also notorious for their wastewater foam problems.

Wastewater Treatment. Defoamers are now used extensively to treat wastewater in many municipal and industrial treatment facilities and also in mining and mineral processing. Benefits are aesthetic, environmental, and economic. Aeration basins are rid of unsightly, troublesome foam, and water and energy are conserved, sometimes markedly, by allowing more efficient use of mechanical equipment. Most foaming problems occur either at the biological treatment step or the effluent discharge step (48). Aeration (qv) is necessary in biological treatments to allow microorganisms to breathe, but the agitation produces foam. Defoamers that do not disrupt this process must be selected. Similarly, in many processes nowadays there is a trend towards greater recycling of treated streams. Some of the upstream processes may depend on foaming and other surface active phenomena such as flotation and flocculation steps, and care must be taken to use defoamers that do not interfere with these processes. Effluent foam discharge is illegal in some countries including the U.S. where the EPA prohibits discharge of floatable materials in the effluent stream (48).

Economic Aspects

Accurate information on the size of the defoamer market is impossible to obtain. There are too many types of materials and suppliers involved. Particularly for the more common oils and surfactants, defoaming is a very small part of their total usage, and no public information is available on what fraction of manufacturers' sales is in the area of foam control. Even for more expensive materials such as the poly(alkylene oxide) block copolymers, there is no way of distinguishing between their use as defoamers and other significant surfactant uses such as de-emulsifiers.

In 1979 the *Encyclopedia* estimate was a market for defoamers of all types greater than 136,000 metric tons per year in the United States, and at least equal to that in the rest of the world. More recent reviews (3,5) have simply repeated this estimate. Some of the principal use areas such as pulp and paper have experienced modest growth in the last decade, and new products, such as powdered antifoams, have found new applications, but in general this is a mature market and improvements in efficiency in some products, and also in the methods of dispersing them, partly compensate for growth in other areas. A reasonable 1990 estimate for U.S. usage is 150,000 to 170,000 metric tons per year with the rest of the world probably now slightly exceeding that amount. No typical selling price for defoamers can be quoted because of the wide diversity of active materials and product types. The cheapest unformulated defoamer types cost as little as \$0.6 kg in bulk quantities in 1990 before the rise in crude oil prices. Some custom-tailored defoamers cost over \$5.0 kg. Usages vary from less than 0.1 kg of defoamer per ton of product to over 2.0 kg per ton of product, and it is not the unit cost of the defoamer that matters but the cost per unit of product produced. It often occurs that the controlled use of a relatively expensive defoamer results in a lower cost per unit of product than use of an inexpensive defoamer.

Test Methods

The ultimate test of a defoamer is an actual field trial, and occasionally this is the only testing carried out. More usually some laboratory-scale evaluation of several different products is conducted before a recommendation is made for a suitable, economical defoamer for a specific application. Although suppliers have their chosen standard foaming surfactants, it is usual to work with the potential customer's foaming medium. Often this work is done at the plant site to obtain fresh foaming liquors. Establishing that a given defoamer is effective in a particular application is only part of the testing required. The absence of any adverse effect on the final product and the manufacturing and use environments must also be determined. In addition, to be marketable the defoamer must be cost-effective and convenient and easy to handle for the customer. A useful account of the practical selection of defoamers has been given (41).

There are many laboratory methods for testing the relative merits of one defoamer against another. It is a simple matter to measure foam height as a function of time to compare the performance of various foam surfactants and defoamers. Unfortunately, this simplicity has led to a wide variety of methods and conditions used with no standard procedure that would make the measurement of foaminess as characteristic of a solution as its surface tension or viscosity. It has been suggested that the time an average bubble remains entrapped in the foam is such a quantity (49), but very few workers in the defoamer industry have adopted this proposal. In practice, a wide variety of methods are used that generally fall into one of five main categories:

- (1) *Pneumatic Methods.* Gas introduction is controlled by injection through capillary tubes, sintered-glass spargers, diffuser stones, and the like.
- (2) *Dynamic Methods.* This is a subdivision of the pneumatic class; foam heights or volumes are monitored while the gas continues to produce bubbles.
- (3) *Shaking Methods.* Agitation is the easiest way of producing foam, but the results are very dependent on the details of the shaking procedure. Foam height, or for more viscous fluids specific gravity changes, are usually measured.
- (4) *Pour Methods.* The liquid is poured or drained from one vessel into another. This approach is best limited to foams produced from dissolved gases in the liquid.
- (5) *Stirring and Blending Methods.* Like shaking methods, these are very dependent on procedural and equipment details, but they are simple to use and are widely employed for comparative purposes.

This diversity of test methods is reflected in ASTM recommendations (50). Pour (D1173-53 (reapproved 1986)), shaking (D3601-88), and blending (D3519-88) methods are suggested for aqueous solutions. Pneumatic methods are recommended for lubricating oils (D892-89) and engine coolants (D1881-86). Despite their omission by ASTM, dynamic methods are probably the most satisfactory for defoamer evaluation (9,43). The foaming solution is usually recirculated through a vertical cylinder where foam heights can be measured as a function of time. In such a device, a steady state can be achieved with a given foamer system and defoamer metered in at selected concentrations. The test enables various important defoamer characteristics to be measured such as the knockdown time, the time taken to collapse a preformed foam, and the persistence or hold-down time, the time taken for the foam to recover to some agreed level such as half its original height (defoamer half-life). Some very simple dynamic apparatus have been reported (13,52) which could be taken to the customer's plant site, but the usual strategy is to carry out simple shaking or stirring tests on site and carry samples back for more rigorous dynamic or pneumatic testing in the laboratory.

Health and Safety Factors

Defoamers are usually added at low bulk concentrations ranging from a few to a thousand parts per million of the foaming medium. Often the health risk posed by such additives is negligible compared to that of the material being defoamed. Such is the case in the defoaming of asphalt (qv) and phosphoric acid. Sometimes a specific defoamer type foaming medium combination presents a particular problem, so the supplier should always be involved in defoamer selection. Examples are the increase in flammability of polyester textiles with free polydimethylsiloxane (13), and the possibility that mineral oil antifoams may contain precursors that form dioxins in bleached pulp after chlorination (46). Health and safety concerns arise primarily in applications in the food and drug industries. Defoamers can be incorporated directly in these products, as in the production of sugar from sugar beets or in the defoaming of fats for frying potato products, and indirectly, as in the manufacture of paper or plastic packaging materials. The sugar beet example illustrates the value of regulation. Before regulating agencies were made aware of the problem, many producers used spent motor oils as defoamers in this process. Nowadays only specifically approved raw materials can be used for sugar production.

U.S. government regulations governing the use of additives such as defoamers in food and drugs are listed in the Code of Federal Regulations. Title 21 contains the rules established by the Food and Drug Administration; Title 40 covers those that are the concern of the Environmental Protection Agency. For example, part 173.340 of Title 21 deals with defoamers that may be safely used in processing foods, whereas Part 180.1001 of Title 40 lists those materials exempt from the requirement of tolerance levels in pesticide chemicals, including defoamers used therein. Other parts of Title 21 that cover defoamer uses include 176.200 on coatings and 176.210 on the manufacture of paper and paperboard. One defoamer is also used as an active drug ingredient—the antifoaming silicone, Dimethicone [8050-81-5]. Such use is regulated by the FDA under part 332 of Title 21. Regulations are subject to change and the CFR is revised at least once each calendar year. It is also kept up-to-date by the individual issues of the *Federal Register*, a daily government publication.

Although the intent may be the same, the details of regulatory practice differ in other countries, and a multitude of regulations exist worldwide. For example, the situation in the former Federal Republic of Germany has been described (3). Under the auspices of the Joint Expert Committee on Food Additives of the FAO-WHO, efforts are being made to establish international guidelines. However, given the changing regulatory climate, the differences in practice between countries, and the wide range and regional differences of antifoam components available, the best advice for those interested in health and safety aspects of defoamers is to contact the producers directly. Their skill and experience are the best defenses against safety hazards associated with the use of defoamers.

BIBLIOGRAPHY

"Defoamers" in *ECT* 3rd ed., Vol. 7, pp. 430–448, by Irwin Lichtman and Ted Gammon, Diamond Shamrock Chemical Co.

1. J. C. Colbert, *Foam and Emulsion Control Agents and Processes: Recent Developments*, Noyes Data Corp., Park Ridge, N.J., 1981.
2. H. T. Kerner, *Foam Control Agents*, Noyes Data Corp., Park Ridge, N.J., 1976.
3. R. Hofer and co-workers, in B. Elvers, J. F. Rounsaville, and G. Schulz, eds., *Ullmann's Encyclopedia of Industrial Chemistry*, Vol. A11, 5th ed., VCH Publishers, New York, 1988, pp. 465–490.
4. J. W. Simmons, R. N. Thornton, and R. J. Wachala, in L. J. Calbo, ed., *Handbook of Coatings Additives*, Marcel Dekker, Inc., New York, 1987, pp. 147–175.
5. M. J. Owen, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, Vol. 2, 2nd ed., John Wiley & Sons, Inc., New York, 1985, pp. 59–72.
6. P. E. Levesque and N. S. Mount, *Plastverarbeiter* **18**, 137 (1967).
7. N. F. Chinoy, S. W. Dean, Jr., and R. Derby, *The Role of Acetylenic Glycols in the Growth of Water Based Coatings*, *Air Products Technical Paper* 120–313, Air Products and Chemicals Inc., Allentown, Pa., 1979.
8. **U.S. Pat. 4,208,385** (June 17, 1980), M. L. Robbins and E. R. Ernst (to Exxon Research and Engineering Co.).
9. **U.S. Pat. 4,082,690** (Apr. 4, 1978), K. W. Farminer (to Dow Corning Corp.).
10. **U.S. Pat. 2,933,382** (Apr. 19, 1960), F. M. Fowkes and F. G. Bollo (to Shell Oil Co.).
11. **U.S. Pat. 3,951,853** (Apr. 20, 1976), D. W. Suwala (to Diamond Shamrock Co.).
12. **U.S. Pat. 3,076,768** (Feb. 5, 1963), F. J. Boylan (to Hercules Powder Co.).
13. R. E. Patterson, *Text, Chem. Color.* **17**(9), 181 (1985).
14. P. R. Garrett, *J. Colloid Interface Sci.* **69**, 107 (1979).
15. J. W. Gibbs, *Collected Works*, Vol. 1, Longman-Green and Co., New York, 1928, p. 300.
16. A. Prins and M. Van den Tempel, *Proc. IVth Int. Congr. Surface Active Substances*, Brussels, **2**, 1119 (1964).
17. I. C. Callaghan and co-workers, *Colloids and Surfaces* **8**, (1983).
18. H. Saito and S. Friberg, *Pramana*, 1st Suppl., 537 (1975).
19. N. L. Jarvis, *J. Phys. Chem.* **70**, 3027 (1966).
20. S. Hard and R. D. Neuman, *J. Colloid Interface Sci.* **120**, 15 (1987).
21. C. B. McKendrick, S. J. Smith, and P. A. Stevenson, *Colloids and Surfaces* **52**, 47 (1991).
22. M. J. Rosen, *Surfactants and Interfacial Phenomena*, John Wiley & Sons, Inc., New York, 1978, p. 164.
23. I. R. Schmolka, in M. J. Schick, ed., *Nonionic Surfactants*, Marcel Dekker, Inc., New York, 1966, p. 309.
24. S. Wu, in J. Brandrup and E. H. Immergut, eds., *Polymer Handbook*, 3rd ed., John Wiley & Sons, Inc., New York, 1989, p. VI 411.
25. M. Sato, *Proc. Japan Acad.* **31**, 713 (1955).
26. Z. Rymuza, *J. Colloid Interface Sci.* **112**, 221 (1986).
27. A. Halpern, *J. Phys. Colloid Chem.* **53**, 895 (1949).
28. A. I. Vogel and P. M. Cowan, *J. Chem. Soc.*, 16 (1943).
29. S. J. Ross, *The Inhibition of Foaming*, *Bulletin No. 63*, Rensselaer Polytechnic Institute, Troy, N.Y., 1950.
30. W. D. Harkins, *J. Chem. Phys.* **9**, 552 (1941).
31. A. Prins, *Paper 539, presented at 59th ACS Colloid and Surface Symposium*, June, 1985.
32. I. A. Lichtman, V. S. Sinka, and D. W. Evans, *Assoc. Mex. Tec. Ind. Cel. Papel. Bol.* **15**, 26 (1975).
33. S. Ross and G. Nishioka, in P. Becher and M. N. Yudenfreund, eds., *Emulsion, Latices and Dispersions*, Marcel Dekker, Inc., New York, 1978, pp. 237–256; S. Ross and G. Nishioka, *J. Colloid Interface Sci.* **65**, 216 (1978).

34. G. C. Frye and J. C. Berg, *J. Colloid Interface Sci.* **127**, 222 (1989).
35. M. P. Aronson, *Langmuir* **2**, 653 (1986).
36. J. W. McBain and co-workers, *Foaming of Aircraft-Engine Oils as a Problem in Colloid Chemistry*, NACA ARR No. 4105, National Advisory Committee for Aeronautics, Stanford, Calif., 1944.
37. N. Dombrowski and R. P. Fraser, *Phil. Trans. Roy. Soc. (London)*, **A247**, 13 (1954).
38. A. Dippenaar, *Int. J. Miner. Process.* **9**, 1 (1982).
39. R. D. Kulkarni, E. D. Goddard, and B. Kanner, *Ind. Eng. Chem. Fundam.* **16**, 472 (1977); R. D. Kulkarni, E. D. Goddard, and B. Kanner, *J. Colloid Interface Sci.* **59**, 468 (1977).
40. G. C. Frye and J. C. Berg, *J. Colloid Interface Sci.* **130**, 54 (1989).
41. J. McGee, *Chem. Eng.* **96**(4), 131 (1989).
42. G. C. Sawicki, *J. Am. Oil Chem. Soc.* **65**, 1013 (1988).
43. W. C. McGregor, J. F. Weaver, and S. P. Tansey, *Biotechnol. Bioeng.* **31**(4), 385 (1988).
44. W. T. Flannigan, *Spec. Chem.* **8**(5), 420 (1988).
45. I. C. Callaghan and co-workers, *Spec. Publ. Roy. Soc. Chem.* **59** (*Ind. Appl. Surfactants*), 48–57 (1988).
46. G. C. Sawicki and J. W. White, *Chemspec Europe '89 BACS Symposium*, Manchester, UK, 1989.
47. **U.S. Pat. 3,912,652** (Oct. 14, 1975), J. A. Colquhoun (to Dow Corning Corp.).
48. D. J. Avery, *Pollut. Eng.* **21**(10), 113 (1989).
49. J. J. Bikerman, *Foams*, Springer-Verlag, New York, 1973.
50. *Annual Book of ASTM Standards*, American Society of Testing and Materials, Easton, Md., 1990.
51. **U.S. Pat. 3,107,519** (Apr. 27, 1960), E. P. McGinn (to Drew Chemical Corp.).
52. J. A. Colquhoun, *Text. Ind. Atlanta* **137**, 100, 103, 105 (1973).

Michael J. Owen
Dow Corning Corporation

DEFORMATION AND RECORDING MEDIA.

See INFORMATION STORAGE MATERIALS.

DEGRADABLE POLYMERS.

See PLASTICS, ENVIRONMENTALLY DEGRADABLE.

DEGREASING AGENTS.

See METAL SURFACE TREATMENT.

DEHUMIDIFICATION.

See AIR CONDITIONING.

DENSITY AND SPECIFIC GRAVITY.

See THERMAL, GRAVIMETRIC, AND VOLUMETRIC ANALYSIS.

DENTAL MATERIALS

Dental ceramics,
Dental cements,
Dental plasters,
Dental investments,
Dental waxes,
Dental alloys and metals,
Polymeric dental materials,
Abrasives,
Therapeutic dental materials,
Calcium phosphate materials,
Dental implants,

Dentistry, as approved by the Federation Dentaire Internationale (FDI) is the "science and art of preventing, diagnosing, and treating diseases and malformations of and injuries to the teeth, mouth, and jaws, and of replacing lost teeth and associated structures." Dental therapy includes the replacement of hard and soft oral tissues lost through disease using inert materials that may be metallic, ceramic, or organic, or composites employing combinations of these three classes. The operative restorations and prostheses are made of amalgam, precious and nonprecious alloys, special cements, synthetic polymers, porcelain, and glass–ceramics (qv) (see CEMENT; GLASS; ENAMELS, PORCELAIN OR VITRIOUS). All must withstand the rigors of the oral environment (see also PROSTHETIC AND BIOMEDICAL DEVICES). The accessory materials needed in the fabrication procedures include synthetic polymers, synthetic and natural gums (qv) and waxes (qv), hydrocolloids, gypsums (see CALCIUM COMPOUNDS: CALCIUM SULFATE), and refractories (qv).

The total value of dental supplies and equipment manufactured in the United States in 1992 is estimated at ca \$1.2 billion. These materials are used by ca 170,000 practicing dentists (1992) and >8,500 commercial dental laboratories employing ca 40–50,000 technicians. The dental materials market is limited because the public does not buy directly except for toothpastes, mouthwashes, toothbrushes, denture aids, etc (see DENTIFRICES).

The dental industry itself is relatively small. It had 504 manufacturing establishments in 1987, only 129 of which employed more than 20 persons. In 1992 there were over 640 dental research projects valued at more than \$106 million and sponsored by government and private organizations in the United States.

Specifications. Standards for dentistry's various unique restorative materials were first developed by Dr. Wilmer Souder and co-workers at the National Bureau of Standards in 1926 with a specification for dental amalgam. In 1928, a cooperative program involving the American Dental Association (ADA) and the Federal Government began. This cooperative dental research program continues at the National Institute of Science and Technology (NIST). The ADA Council on Dental Materials, Instruments, and Equipment is responsible for the development and promulgation of ADA specifications for dental materials and devices. The ADA serves as the secretariat and sponsor for Accredited Standards Committee (ASC) MD156, Dentistry, accredited by the American National Standards Institute (ANSI) which develops the standards. American Dental Association specifications are submitted to ANSI for approval as an American National Standard, following which they are referred to as ANSI ADA Standards. A list of the current standards is available (1).

The formulation of new specifications and the revision or reaffirmation of existing specifications are the responsibility of American National Standards Committee MD156 of ANSI for Dental Materials and Devices. This committee coordinates its work closely with Technical Committee 106,

Dentistry of the ISO.

A primary responsibility of the Food and Drug Administration (FDA) is the enforcement of the Federal Food, Drug, and Cosmetic Act of 1938 and its various amendments, eg, May, 1976, in which dental materials, instruments, and equipment are included. Premarketing clearance requirements apply for establishing the safety and effectiveness of new products. There is a close liaison between the FDA and the ADA standards and certification programs.

The ADA maintains a list of certified dental materials and devices based upon the certification by the maker that the item complies with ADA specification and that the testing for specification compliance of the item is procured in Association laboratories. The ADA also maintains a list of classified dental materials and devices which prove to be acceptable or provisionally acceptable to the Association based upon data submitted by the applicant and data available in the literature.

Biocompatibility. Many of the dental materials are used in contact with body tissues, eg, tooth, bone, soft tissue; and body fluids, eg, saliva and blood; usually for extended periods of time, ie, years. As knowledge of the toxic, allergic, and carcinogenic responses to substances foreign to the body has increased, the need for evaluation of biological responses to materials used in specific locations and for specific functions has become clear. In order to address the need for evaluation, ADA ANSI specification #41, Biological Evaluation of Dental Materials, has been developed. This specification is revised and updated regularly. For materials for which an ANSI ADA specification does not yet exist, the demonstration of acceptable tolerance by specification #41 helps provide to the patient the highest level of protection against injury by the new material.

Dental Ceramics

Ceramic materials are well suited for biomedical applications requiring tissue compatibility, compressive strength, durability, radiopacity, and inertness towards sterilization chemicals and conditions (see CERAMICS). In addition, glass-matrix ceramics, containing appropriate crystalline fillers and colorants, can richly mimic many of the aesthetic and optical properties of natural teeth (2) (see COLORANTS FOR CERAMICS). Limitations to the use of ceramics include design restrictions based upon the often low and variable tensile strengths of ceramic products; susceptibility to stress corrosion, eg, crack extension in the presence of water or other low molecular weight species; and shrinkage during processing.

The principal use of ceramics in dentistry is for the aesthetic restoration of missing teeth or tooth structure, ie, crowns and bridges, primarily as amorphous (glassy) ceramic coatings fused to an underlying metal framework (3). Additional aesthetic use is made of ceramics not supported by metal substructures. These uses include single-tooth crowns, ie, restoring all the external surfaces of a tooth; ceramic fillings, eg, inlays and onlays; veneers, ie, replacement of only the visible surface of anterior teeth; denture teeth; inserts, ie, the substance placed inside of polymer-based filling materials to mitigate against polymerization shrinkage and improve physical properties; and orthodontic brackets. Some semi-aesthetic, generally more highly crystalline, ceramics are utilized as core materials in place of cast metal substructures, to be veneered with more aesthetic ceramics. Nonaesthetic dental ceramics include: implants, ie, single-tooth replacements implanted in bone; biologically active ceramics, based upon calcium- and phosphorous-containing minerals or glasses, used as coatings on dental implants or placed into tooth extraction sites and surgery sites to enhance the maintenance or regeneration of bone; dental cements based upon reactions of acids, organic or mineral, with glass or metal oxide powders; and gypsum products. These nonaesthetic uses of ceramics are covered separately.

Aesthetic dental ceramics are essentially glass-matrix materials with varying volume fractions of crystalline fillers. Crystalline fillers are used in the glass matrix both for dispersion strengthening, usually at volume fractions of 40–70%^o, and for altering optical properties, usually at low volume fractions. Dental ceramics are generally manufactured from two distinct classes of materials, ie, beneficiated feldspathic minerals and glass–ceramics.

Feldspathic Matrix Ceramics. The majority of aesthetic ceramics are based on mined feldspar minerals such as potassium feldspar, eg, orthoclase [12251-44-4] and microcline [12251-43-3], K₂O·Al₂O₃·6SiO₂; sodium feldspar, eg, albite [12244-10-9], Na₂O·Al₂O₃·6SiO₂; and nepheline syenite [37244-96-5], that contains approximately 50%^o albite, 25%^o nepheline, Na₂O·Al₂O₃·2SiO₂, and 25%^o microcline. These feldspars are processed to remove iron-bearing impurities, melted with various additives, quenched in water, and ground to yield powders (frits) of alkali-modified aluminosilicate glasses containing little or no crystalline filler (4,5) (see also GLASS).

Natural tooth colors are developed in these formulations by the addition of color frits, ie, vitreous powders to which colored and or opaquing pigments (qv) have been added. Because aluminosilicate glasses are very aggressive solvents at elevated temperatures, pigments must have inertness toward dissolution as well as color stability at elevated temperatures. Crystalline minerals in the spinel family have been found to be successful pigments (5). Spinel has the general formula MOM'₂O₃, where the divalent metal, M, can be Mg, Fe, Cu, Zn, Ni, Co, or Mn; and the trivalent metal, M', can be Al, Fe, Mn, Ga, or Cr. Particles of metal oxides may also be used as color pigments, eg, the oxides of Ni, Ti, Co, Cu, Cr, Mn, or Sn (6).

Additional raw materials, such as kaolin clay [1332-58-7] and quartz [14808-60-7], may be used as processing aids for factory-based processing, eg, of denture teeth. Kaolin clays form plastic masses useful for molding operations, add Al³⁺ to the glass composition, and aid in sintering because of extremely fine particle size. Quartz can contribute to the shape–stability of vitreous bodies during firing by decreasing flow, as well as add certain aesthetic qualities.

Ceramics designed to be fused to metals are derived from either high potassium feldspars, or from compositions to which potassium oxide [12136-45-7], K₂O, has been added to above approximately 11 mol %^o. This places these ceramics into a phase region from which feldspars melt incongruently, ie, above 1150°C, to form leucite [1302-34-7], K₂O·Al₂O₃·4SiO₂, plus liquid (7). Leucite has a high thermal expansion coefficient and undergoes a displacive transformation, tetragonal to cubic, during heating. This raises the effective coefficient of thermal expansion of the leucite-filled glass into a range compatible with that of cast dental alloys (8,9).

Processing. Traditional ceramic dental restorations are custom fabricated by skilled technicians in specially equipped laboratories. Processing generally involves formation of powder-water slurries that are incrementally built up in layers of varying color and translucency using brushes or spatulas. Consolidation of the greenware is achieved by a combination of vibration and wicking away excess water. Sintering is generally performed at 900–1000°C under vacuum conditions of ca 3.3 kPa (25 Torr) using heating rates in the range of 40–60°C min and hold times of 0.5–10 min at the maximum temperature.

Computer-aided design and manufacturing (CAD CAM) fabrication of customized dental restorations is becoming an increasingly viable alternative, especially for simple restorations (10). Restorations are machined from standardized blocks of completely sintered and internally colored ceramic, either feldspathic or glass–ceramic, (11,12). Feldspathic blocks for CAD CAM processing are formed using extrusion molding of powders utilizing organic binders, as opposed to kaolin clays.

More recently, transfer-molding has been used as a technique to form both greenware and finished ceramic parts (13,14). Mold-making usually follows traditional dental lost-wax techniques used for investment casting. Ceramic greenware has been formed from ceramic compositions incorporating polymeric binders and plasticizers (qv), usually by pressing into a mold at moderately elevated temperatures (14). Finished ceramic parts can be fabricated by pressing filled-glass compositions, designed to have appropriate viscosities and chemical stability at elevated (1100°C) temperatures, into phosphate-bonded molds (13). These transfer-molded dental ceramics are usually product-specific and rely upon unique instrumentation dedicated to each proprietary product.

Strengthening. Strengthening methods for ceramic restorations involve structural approaches, eg, the veneering of metal or high strength ceramic cores, as well as microstructural alterations and surface treatments of the ceramic. Microstructural alteration generally involves increasing the volume fraction of crystalline material in the glass, either by direct addition or by devitrification (ceramming). Surface treatments are used to develop compressive stresses on exterior surfaces by either thermal or chemical routes.

The addition of alumina [1344-28-1], Al₂O₃, to a feldspathic dental glass results in a strengthened ceramic material commonly known as aluminous porcelain (15). The alumina acts as a dispersion strengthening phase. It also decreases the translucency of the sintered filled-glass as a result of differences in refractive index between alumina and feldspathic glasses, and the development of residual porosity. Traditional dental laboratory processing techniques and optical property considerations limit the particle size (90% > 10 μm) and weight percent (40–50 wt%^o) of Al₂O₃ used. A much more highly loaded (>70 wt%^o), dispersion strengthened, alumina-filled ceramic has been developed for use as a core material (16). In this technique a much finer alumina (80% > 2 micrometers) is slip-cast onto a refractory material and initially sintered into a highly porous form. A high index of refraction (lanthanum modified)

feldspathic glass is infiltrated into the porous alumina preform, yielding a translucent composite ceramic having strength properties approaching that of polycrystalline alumina (see COMPOSITE MATERIALS, CERAMIC MATRIX).

Leucite has been used for dispersion strengthening both in products provided as powders to be processed via traditional water-slurry techniques and in formulations for high-temperature transfer molding (13,17). Because of its index of refraction, leucite is aesthetically superior to alumina as a filler for feldspathic glasses, but it is not superior as a dispersion strengthening filler. Leucite generally is introduced into dental ceramics via nucleation and growth processes from high potassium-content aluminosilicate glasses.

Ion-exchange approaches and thermal tempering have been evaluated for strengthening dental ceramics (18,19). Both of these approaches are aimed at placing external surfaces of dental ceramic restoration in compression. Only ion exchange is promoted commercially and is not in extensive use.

Dental Glass–Ceramics. Glass–ceramics (qv) are highly crystalline ceramics having some residual glass matrix prepared by the controlled crystallization of glasses (20). Crystalline phase(s) are formed from elements within the parent glasses by subjecting them to carefully regulated thermal treatments, called ceramming. Generally, ceramming treatments involve separate control, at two different temperatures, over both the nucleation and growth stages of crystallization within the glass.

Glass–ceramics available for restorative dentistry are based on three different crystalline phases, ie, tetrasilicic fluoromica, $\text{KMg}_{2.5}\text{Si}_4\text{O}_{10}\text{F}_2$, β -quartz crystals, and leucite (13,21,22). Fluoromica develop with a plate-like morphology having a relatively low-energy cleavage path along their basal plane. These tendencies towards basal-plane cleavage and the interlocking of crystals at high volume fraction provide an effective strengthening mechanism (23). Fluoromica crystals and the residual glass matrix have relatively similar indices of refraction.

The β -quartz solid solution phase is precipitated from a lithium aluminosilicate glass. This ceramming step must be carefully controlled in order to avoid conversion of the β -quartz to β -spodumene which would rapidly opacify the ceramic. Internal coloration is achieved by interactions between Ce and Ti species in the residual glass. One aesthetic leucite-containing glass–ceramic is manufactured from a glass melt (composition in wt %) of 63% SiO_2 , 17.7% Al_2O_3 , 11.2% K_2O , 4.6% Na_2O , 0.6% B_2O_3 , 0.4% CeO_2 , 1.6% CaO , 0.7% BaO , and 0.2% TiO_2 (13).

Processing. Glass–ceramics are formable into complex shapes while still in the glassy state and essentially retain their finished dimensions during ceramming. Glass for the fluoromica-containing ceramic is cast into a phosphate-bonded refractory mold using a special high temperature centrifugal casting machine at temperatures between 1350°C and 1365°C (21). Following formation, the glass castings are cerammed at a temperature of 1070°C for 6 hours. Colored feldspathic glasses are fired in very thin external layers in order to develop proper dental aesthetics.

Glass of the β -quartz-containing ceramic is melted at 1600°C and pressure-cast, using metal molds, into a variety of small shapes suitable for insertion into teeth being filled with polymer-based restorative materials. Appropriate tooth color, index of refraction, and improved physical properties are developed during a ceramming treatment (22). The surface of the glass ceramic parts is treated using an organo-functional silane to enhance bonding between the glass–ceramic and restorative dental resins.

Parent glass for one leucite-containing glass–ceramic is melted at 1450°C, cast as simple-shaped billets, and cerammed by the manufacturer (13). The precerammed billets, already having appropriate tooth color, are loaded into a dedicated high temperature press and transfer-molded at 1100°C into phosphate-bonded molds.

Dental Cements

Dental cements are composites, ie, they have a continuous or matrix phase linked to a discontinuous or reinforcing phase by an interphase. Based on their chemistry, dental cements can be divided into acid–base cements, acrylic or resin cements, and resin-modified acid-base cements, ie, hybrid cement composites. Acid–base cements are either aqueous-based cements or nonaqueous-based cements, although small amounts of water or protoic agents are essential to the acid-base setting mechanism of the latter. Acrylic cements are resin-based composites modified for cement applications. The monomer(s) used in these cements can be simply methyl methacrylate [80-62-6] or various types of multifunctional monomers, eg, BIS-GMA [1565-94-2], triethylene glycol dimethacrylate, [2351-42-0], etc. The setting mechanism is by free-radical addition polymerization, either chemical, photochemical, or a combination of the two. The hybrid cement-composites include a vinyl resin system as well as the usual acid–base components of the cement. They have a dual, ie, ionic and free radical, setting mechanism.

Although used in relatively small quantities, dental cements are essential in a number of dental applications as temporary, intermediate and, in some cases, more permanent restoratives; cavity liners and bases; luting agents to bond preformed restorations and orthodontic devices; pulp capping agents and endodontic sealers; components in periodontal packs; and impression pastes. Because this wide range of dental uses makes it virtually impossible for one type of cement to have all the necessary properties demanded, tailored dental cements exist to meet certain specific objectives. Generally, cements are sought that excel in strength, oral environmental resistance, durability, adhesiveness, and biocompatibility.

The oldest acid–base cements, ie, zinc phosphate, and zinc oxide eugenol (ZOE), utilize zinc oxide as the principal base component and orthophosphoric acid or eugenol as the acid component, respectfully. As seen in Table 1, acid–base cements represent an important category of dental cements, although resin-modified glass–ionomers and new types of resin cements are replacing the former types in many dental applications. Research is focused on the improvement of the polyelectrolyte cements, eg, glass–ionomer, and the development of durable, nonshrinking resin or polymer-based adhesive materials with improved biocompatibility.

Table 1. Classification and Composition of Aqueous Dental Cements

Dental cement	Components		Set cement		Uses
	Aqueous acid	Powder base	Material ^a	Amorphous matrix	
zinc phosphate	orthophosphoric acid (Al^{3+} and Zn^{2+}) ^b	calcined zinc oxide, calcined magnesium oxide ^c	oxides	phosphate	luting agent, base
zinc polycarboxylate	poly(alkenoic acids) ^d	zinc oxide and magnesium oxide ^e or stannic oxide	oxides	polycarboxylate	adhesive liner, base, luting agent
glass–ionomer	poly(alkenoic acid)	ion leachable calcium alumino–fluoro-silicate glasses	glass	polycarboxylate	{
metal-modified	poly(alkenoic acid)	calcium alumino–fluorosilicate glass containing fused particulate metal	metallized glass	polycarboxylate	
resin-modified	monomer system ^f , components of initiator system	poly(alkenoic acid), calcium alumino–fluorosilicate glass, component of initiator system	glass	polycarboxylate	
					restorative for cervical lesions, adhesion base, liner luting agent, sealant core buildup materials, sealants

- ^a Partially reacted and unreacted in an amorphous matrix.
- ^b Minor amounts of these ions.
- ^c Present in lesser amounts.
- ^d For example, poly(acrylic acid).
- ^e For example, 9:1.
- ^f For example, 2-hydroxyethyl methacrylate.

Aqueous-Based Cements.

Zinc Phosphate Cements. Zinc phosphate cements are the oldest of the aqueous-based cements (see Table 1) and are still used in a wide range of applications; eg, cavity bases, temporary restoratives, and for the fixation of inlays, crowns, fixed partial dentures (bridges), posts, facings, and orthodontic bands.

In contrast to glass-ionomer cements, zinc phosphate cements are neither as sensitive to the gain or loss of water during setting nor adhesive to tooth structure. Retention is done by mechanically interlocking the irregularities of surfaces being cemented. The setting times in the mouth are 3–11 min, typically 7–8 min. Dimensional change upon setting may be less than 0.1% shrinkage, if the mass is protected from water loss, and the compressive strength at 24 h is 88–143 MPa (13,000–20,700 psi). The solubility of the set cement in distilled water is generally from 0.1–0.2% for specimens exposed for 7 days.

The powder consists of calcined zinc oxide [1314-13-7] and magnesium oxide [1314-13-2] in ratio of ca 9:1. The liquids are usually concentrated (45–63 wt %) phosphoric acid solutions buffered by aluminum salts and, in some instances, by both aluminum and zinc salts (see PHOSPHORIC ACIDS AND PHOSPHATES) (24). The compounds formed by reaction of the powder and liquid are predominately noncrystalline phosphates of zinc [7779-90-0], magnesium [7757-86-0], and aluminum [7784-30-7] (25–27). The set cement consists of unreacted cores of powder particles bound together in a generally structureless gel matrix of phosphates of zinc, magnesium, and aluminum. The proportion of unused powder particles and phosphate matrix in the set cement varies with the amount of powder incorporated into a given amount of liquid to produce the optimum consistency for the specific use. The set cement, having a minimum amount of gel matrix and maximum amount of particles, has the best physical properties for use in the mouth.

Except for the little used silicate cements, which use calcium aluminofluorosilicate glasses in place of zinc oxide–magnesium oxide base powder, modifications of the zinc phosphate cement have been minimal (28). A novel modification of the zinc phosphate cement is the inclusion of phosphate-based organic polymers in the formulation. This improves both its strength and adhesion to dentin (29). The most important classes of new acid–base cements are those based on poly(alkenoic acids), ie, the polycarboxylate cements, the glass-ionomer cements, and the resin-modified glass-ionomer cements (see IONOMERS).

Polycarboxylate Cements. Polycarboxylate cements (30,31) are made by mixing a zinc oxide-based powder and an aqueous solution of poly(acrylic acid) [9003-01-4] or similar polyacid (see ACRYLIC ACID). The biological effects of these cements on soft and mineralized tissues are mild (32). The effect on the pulp is similar to that of the zinc eugenolate cements (33). This type of cement seems to have overall excellent biocompatibility (28). When freshly mixed, the carboxylic acid groups convert to carboxylates, which seems to signify chemical adhesion mainly via the calcium of the hydroxyapatite phase of tooth structure (32,34–39). The adhesion to dentin is reduced because there is less mineral available in this substrate, but bonding can be enhanced by the use of mineralizing solutions (35–38). Polycarboxylate cement also adheres to stainless steel and clean alloys based on multivalent metals, but not to dental porcelain, resin-based materials, or gold alloys (28,40). It has been shown that basic calcium phosphate powders, eg, tetracalcium phosphate [1306-01-0], Ca₄(PO₄)₂O, can be substituted for zinc oxide to form strong, hydrolytically stable cements from aqueous solution of polyacids (41,42).

The compressive strength of polycarboxylate cements at cementing consistency is 55–85 MPa (8,000–12,000 psi). Typical diametral tensile strength ranges from 8–12 MPa (1160–1740 psi). The solubility and disintegration in distilled water after 7 days at 37°C is 0.04–0.08 wt %, and is not reflected in clinical performance.

The powder contains zinc oxide and magnesium oxide (36), and the liquid contains an aqueous solution of an acrylic polycarboxylic acid. Water settable cements have been formulated by inclusion of the solid polyacid in the powdered base component. The set cement mainly consists of partially reacted and unreacted zinc oxides in an amorphous polycarboxylate matrix (27,28).

Glass-Ionomer Cement. The glass-ionomer polyelectrolyte system was developed primarily as a restorative for anterior teeth and erosion cavities; a general cement; a cavity liner; and a base, pit, and fissure sealant (27,43–48).

The glass-ionomer cement has some translucency in contrast to that of the opaque polycarboxylate cement. The setting time is approximately 2–4 minutes. Compressive strength in 24 h is 154–175 MPa, and diametral tensile strength is 7–19 MPa (1000–2700 psi). Solubility in distilled water after 7 days is ca 0.44%. Glass-ionomers used as restoratives are available in several tooth colored shades. They lack the strength necessary for stress-bearing restorations, but the fluoride from the aluminofluorosilicate glass is slowly released from the filling, possibly inhibiting recurrent caries (49).

Glass-ionomer cement is based on the reaction between calcium aluminofluorosilicate glass powders and aqueous solutions of polymers and copolymers of acrylic acid. Setting modifiers are included to improve working and hardening characteristics (50–52). Tartaric acid [87-69-4] is particularly effective in both prolonging the working time and sharpening the setting time (52). Water settable versions of this cement also are available (53).

These poly(alkenoic acid) based cements are relatively high modulus cements that adhere to enamel and to dentin with minimal surface preparation, and are rapidly replacing the more traditional dental cements in many applications. As with the zinc polycarboxylate cements, adhesive strength to enamel is higher than to dentin, presumably because of the lower mineral content of the latter. But, the cement is weakly adhesive to either substrate because of the low tensile, shear, and flexural strength inherent in this brittle material (54–57). Using a sandwich or laminate technique, the glass-ionomer cement has been successfully used to bond composite resin restorations to dentin (58–60).

Excessive hydration, during early stages of the setting reaction, adversely affects the properties of the glass-ionomer cement. Faster setting cements have been developed in an attempt to correct this deficiency. Moreover, the hardened cement has a propensity to dehydration, even after six months; proper maintenance of a water balance is critical not only during the initial setting but also during maturation of the cement. A tendency toward brittle, even catastrophic failure, and severe erosion under acidic conditions are the main deficiencies of this cement (61). The introduction of light cured glass-ionomer cements has eliminated the need for protection against water gain or loss during setting and alleviated the erosion problem.

One approach for ameliorating the highly brittle nature of these cements has involved the use of tougher, more ductile fillers (62,63). Another approach for improving the overall properties of traditional glass-ionomer cements involves the development of hybrid cement-composites and resin-modified cements (64–68).

Metal-Modified Glass-Ionomer Cements. In an attempt to improve the physical strength of the material, metal-modified glass-ionomer cements were developed from metallized glass powders called cermets (69,70), prepared by the sintering of soft ductile metals, eg, silver or gold, onto ion leachable particulate glass (63,71). The glass-ionomer cermet cements exhibit improved abrasive wear, but the overall strength properties and tendency to brittle failure are not significantly improved.

The glass-cermet cements contain a base powder of modified leachable glass, eg, calcium aluminofluorosilicate glass that contains about 40% by weight of fused silver. The polyacid components, the chemistry of setting, and most mechanical properties are the same as for conventional glass-ionomers. Compared to conventional glass-ionomer cements, the silver cermet cement shows enhanced resistance to abrasive wear (72). Originally designed as a replacement for amalgam fillings, the cermet cements lack the strength, toughness, and durability of this metallic restorative. The silver cermet cement is recommended as a core buildup material.

Resin-Modified Glass-Ionomer Cements. Resin-modified glass-ionomer cements are based on poly(alkeonic acid) systems that have

been modified to contain vinyl polymerizable groups and or compatible monomers (64–68,73). The mechanism of setting involves both free radical vinyl addition polymerization as well as the ionic acid-base reactions of the conventional glass–ionomers (64–68,73).

Resin-modified glass–ionomer lining and restorative materials add a multifunctional acidic monomer to the poly(acrylic acid) [9003-01-4] liquid component of the system. Once the glass powder and liquid are mixed, setting can proceed by the acid–glass–ionomer reaction or the added monomer can be polymerized by a free-radical mechanism to rapidly fix the material in place (74,75). The cured material still retains the fluoride releasing capabilities of a glass–ionomer.

Polymer-modified glass–ionomers also have been prepared and involved only acid-base setting mechanisms (67). A related cement contains both compatible polymerizable monomer systems and polymers along with the conventional glass–ionomer ingredients (67). These hybrid cements have mechanical properties similar to those of conventional glass–ionomer cements, but because of their tougher nature are less likely to undergo catastrophic failure.

Polymeric Calcium Phosphate Cements. Aqueous solutions of polymers such as poly(acrylic acid), poly(vinyl alcohol), gelatin, etc, and or autopolymerizable monomer systems, eg, 2-hydroxyethyl methacrylate, glycerol dimethacrylate, calcium dimethacrylate, etc, have been used as liquid vehicles (41,42,76) for the self-setting calcium phosphate cement derived from tetracalcium phosphate and dicalcium phosphate [7757-93-9].

Nonaqueous-Based Cements. For many years dentistry utilized a number of nonaqueous products based on the interaction of phenolic chelating agents and similar compounds and a metal oxide or salt (see Table 2). The most exploited system, dating back to 1884, has been the zinc oxide–eugenol cement. Many modifications and improvements have been suggested, such as the use of rosin (77), hydrogenated rosin (78), polymers such as polystyrene (53), and various accelerating systems (79,80). In 1955 a critical study of the basic reaction showed that zinc oxide forms chelate compounds with many materials that have a structure similar to eugenol (81), ie, a methoxy group ortho to a phenolic hydroxyl group. Carboxylic acids and nonphenolic chelating agents also have been investigated as the acid component of several noneugenol cements (82).

Table 2. Classification and Composition of Nonaqueous Dental Cements

Dental cement	Components		Set cement		Uses
	Liquid	Powder base	Material	Matrix	
zinc oxide–eugenol (ZOE)	eugenol	zinc oxide ^a	zinc oxide ^b	zinc eugenolate	temporary luting and filling material; pulp capping; periodontal packs
2-thoxybenzoic acid (EBA)–eugenol	2-ethoxybenzoic acid, eugenol	zinc oxide ^a	zinc oxide ^b	complex zinc eugenolate–EBA	linear and base, temporary filling; temporary cementation
calcium chelate cement	phenolic esters of salicylate acid	calcium hydroxide plus fillers	calcium hydroxide ^c	calcium salicylate	liner and base
resin cements	acrylic monomers ^d	polymeric and or mineral fillers		reinforced acrylic polymeric	cementation, sealants

^a Reinforced cements contain other fillings that may be reactive.
^b Partially and unreacted; also other mineral or polymeric fillers present.
^c Partially reacted; and or fillers present.
^d Newer types employ surface active resins.

Zinc Oxide–Eugenol Cements. Zinc oxide–eugenol cements have many uses in dentistry. The admixture of powdered zinc oxide [1314-13-2] and liquid eugenol [97-53-0], C₁₀H₁₂O₂, forms a bland, easily mixed paste having excellent working time but slow-setting characteristics. Accelerators can be added to the cement to decrease the set time. Compositions of the basic oxide–eugenol cement can be easily adjusted to serve a variety of dental needs. Some of these compositions are of a cementing nature, but many are useful in other applications, eg, impression pastes and surgical packs. Other uses of the zinc oxide–eugenol cements include temporary fillings, root canal fillings, linings for deep-seated cavities under restorations, the cementing of temporary or semipermanent acrylic crown forms or splinted crowns on prepared tooth stumps, and sealing-in treatments.

Suitably formulated, zinc oxide–eugenol cements can be designed having viscosities from thin fluid mixes to heavy puttylike masses. The setting times can be varied from minutes to several hours. Crushing strengths of up to 34 MPa (4930 psi) after 24 h have been reported. The degree of hardness or plasticity of the final set mass can vary from a hard friable mass to a hard strong mass to a soft gummy mass, according to the conditions required. The cements show excellent dimensional stability upon aging. Shrinkages as low as 0.2° after 48 h have been measured. Long-term disintegration (20–22 wk) measurements on conventional zinc oxide–eugenol materials indicate a continual leaching of the eugenol. This loss of eugenol leads to a matrix breakdown and a marked reduction in mechanical properties. Reinforcing agents have been used to enhance mechanical properties (83,84).

The cementing applications employ a variety of products compounded either as powder–liquid two-part formulations, or as two-part paste-type products. The zinc oxide portion of the formula may contain modifying agents to improve the physical properties of the set paste, or to act as accelerating agents to control the setting rate of the mix. Accelerators, such as zinc acetate [557-34-6] and zinc stearate [557-05-1], and reinforcing agents, eg, rosin [8050-09-7], polymerized rosin, hydrogenated rosin, polymers, calcium phosphate [7758-23-8], magnesium oxide [1309-48-4], quartz [14808-60-7], and aluminum oxide [1344-28-1] (85), have been used in the basic cement formulations. Many other additives, including medicinal ones, have also been employed (86).

The eugenol portion of the formula may consist of eugenol alone or may be compounded into a viscous fluid by the addition of various additives. The addition of rosin; polymerized rosin; hydrogenated rosin; plasticizing oils, eg, olive oil; accelerators or retarders; medicinal additives; and other modifiers serve to extend the usefulness of the basic system.

Zinc Oxide–Eugenol Impression Pastes. These two-paste systems are usually supplied in collapsible tubes; one tube contains the white zinc oxide and the other contains the brown eugenol. Approximately equal amounts of the two pastes are mixed to form a smooth, homogeneous mass. The impression paste is spread thinly over the surface of a custom-made tray. The coated tray is positioned in the mouth and held motionless until the paste mix hardens, ca 2–5 minutes. The fluid consistency of the impression paste allows the wash to faithfully copy the finest detail of mucosal surface contours. Impression pastes can have differing characteristics of mix consistency, setting rate, and set-paste hardness. Hard-setting and soft-setting pastes are recognized in ADA specifications.

The zinc oxide–eugenol impression pastes produce a rigid impression with good accuracy and good reproduction of surface detail. They have sufficient resistance so that borders can be built up if the tray is slightly deficient in any area. They set to a cement-like hardness, and the resulting impression can be taken in and out of the mouth repeatedly, giving an opportunity to test for stability and tissue adaptation. In addition they register detail well, and are quite stable dimensionally.

The compositions of zinc oxide–eugenol impression pastes are similar to those of the zinc oxide–eugenol cements (86). Variations in specific characteristics are achieved by the proportions of the ingredients (87). Properties vary in commercial products (88). The modifications of the zinc oxide–eugenol system intended for bite-registration pastes may include agents to increase the body or thixotropic character of the unset mix to improve

the handling and utility for this specialized use.

Zinc oxide–eugenol impression pastes are used primarily as corrective washes over compound impressions, as veneer impressions, as temporary liners or stabilizers in base-plates and dentures, and as bite-registration pastes for recording occlusal relationships in inlay, crown, and bridge techniques. The test methods for zinc oxide impression pastes are outlined in ADA specification no. 16 and Federal specification U-I-500.

Surgical Pastes. Zinc oxide–eugenol pastes having essentially the same formulation as impression pastes are used as dressing for gingival protection and healing following periodontal surgery. These pastes have delayed setting times, are less brittle and are weaker after hardening than the impression pastes.

Some mouth tissues are sensitive to eugenol, and a chronic gastric disturbance may occur in some patients after wearing a surgical pack for 2–3 weeks. A noneugenol material similar to zinc oxide–eugenol can be formed by a reaction between zinc oxide and some carboxylic acids. Carboxylic acids, such as *o*-ethoxybenzoic acid, have been employed in the development of noneugenol cements. Bactericides and other medicines can be added without interfering with the reaction. Because of the ascendancy of elastomeric impression materials, research in the zinc oxide–eugenol impression paste has been minimal (89). There is no specification for the surgical paste.

Zinc Oxide–Eugenol/2-Ethoxybenzoic Acid (EBA) Cements. In an attempt to develop an improved zinc oxide eugenol (ZOE) cement, the reaction of zinc and other oxides with other liquid chelating agents as well as carboxylic acids has been investigated. The only ones of commercial importance have been those based on 2-ethoxybenzoic acid [134-112] and eugenol, ie, EBA cements (85,90–92). Composition parameters are critical in maximizing strength properties and minimizing solubility.

This type of cement has been further improved by the substitution of *n*-hexyl vanillate [84375-71-3] and similar esters of vanillic acid [121-34-6] and or syringic acid [530-57-4] for eugenol (93–95). These substituted cements are strong, resistant to dissolution, and, unlike ZOE and EBA cements, do not inhibit the polymerization of resin-base materials. Noneugenol cements based on the acid–base reaction of zinc and similar oxides with carboxylic acids have been investigated, and several promising types have been developed based on dimer and trimer acids (82).

Calcium Chelates (Salicylates). Several successful dental cements which use the formation of a calcium chelate system (96) were developed based on the reaction of calcium hydroxide [1305-62-0] and various phenolic esters of salicylic acid [69-72-7]. The calcium salicylate [824-35-1] system offers certain advantages over the more widely used zinc oxide–eugenol system. These products are completely bland, antibacterial (97), facilitate the formation of reparative dentin, and do not retard the free-radical polymerization reaction of acrylic monomer systems. The principal deficiencies of this type of cement are its relatively high solubility, relatively low strength, and low modulus. Less soluble and higher strength calcium-based cements based on dimer and trimer acid have been reported (82).

The formulation of calcium chelate materials is based upon the formation of a low-solubility chelate between calcium hydroxide and a salicylate. Dycal utilizes the reaction product of a polyhydric compound and salicylic acid. Other salicylic acid esters can be similarly used. Vehicles used to carry the calcium hydroxide, extenders, and fillers may include mineral oil, *N*-ethyl-*p*-toluenesulfonamide [80-39-7], and polymeric fluids. The filler additions may include titanium dioxide [13463-67-7], zinc oxide, silica [7631-86-9], calcium sulfate, and barium sulfate [7727-43-7]. Zinc oxide and barium sulfate are useful as x-ray opacifying agents to ensure a density greater than that of normal tooth structure. Resins, rosin, limed rosins, and modified rosins may serve as modifiers of the physical characteristics in both the unset and set states.

The physical characteristics of the calcium salicylate system compare favorably with those of the zinc oxide–eugenol products. The setting times are subject to modification over a wide range. However, they are somewhat more critical to control than those of the zinc oxide–eugenol system owing to the greater solubility of calcium hydroxide as compared with that of zinc oxide. Compressive strengths in excess of 14 MPa (2030 psi) have been obtained in 24 hours. The set product may be formulated to have a hard, brittle, resinous fracture or may be modified to a tough, strong, semiplastic mass. The dimensional stability within a 48 h period varies somewhat with the basic system. Shrinkages of only 0.2° in 48 h have been obtained, but 0.4° may be reached in some compositions. Solubilities, when tested according to ADA specification no. 8 for zinc phosphate cements, range from <1 to 14° for the formulations based on polyhydric esters of salicylic acid. The calcium salicylate cements are not retarders for the free radical polymerization of resin-based materials, eg, composites. A visible light activated calcium hydroxide cement (98,99) has been developed.

Calcium hydroxide has been recognized as a desirable lining for deep-seated cavity preparations involving pulp exposures, or near exposures. It stimulates the formation of secondary dentin that forms a natural pulp-protecting wall. Water slurries or mucilage suspensions of calcium hydroxide provides the therapeutic benefits of the material but with no mechanical strength. Cements of hard-setting calcium hydroxide-containing compositions compare favorably with the better qualities of zinc oxide–eugenol cements. The former provide both the therapeutic values of calcium hydroxide and the structurally strong bases needed for permanent restorations of metal, porcelain, plastic, or silicate cement.

The calcium chelate cements are limited to the use of a cavity liner. They may be placed directly over an exposed tooth pulp to protect the pulp and stimulate the growth of secondary dentin, or used as a therapeutic insulating base under permanent restorations. The high alkalinity and high solubility of these materials prohibits use in close proximity to soft tissues or in contact with oral fluids.

The chelated calcium cementing materials are supplied as two-part paste products. In use, equal parts of the two pastes are thoroughly mixed together to give a fluid mass that can be applied without pressure over an exposed tooth pulp or in a deep-seated cavity. Under the influence of the oral temperature and humidity, the fluid mass sets to a hard, strong, therapeutic protective seal.

Resin Cements. Early resin cements formed by free radical polymerization, available since 1952, were based primarily on poly(methyl methacrylate) [9011-14-7] and its monomer (100). Relatively rapid setting times, high polymerization shrinkage, and difficulty of removing excess hardened cement from the interproximal spaces and from beneath free gingival margins are principal deficiencies of this type of resin cement (see METHACRYLIC POLYMERS). More recent resin cements are based on high molecular weight, fluid, di- and multifunctional methacrylates, some containing relatively high loadings of fillers, such as fine glass powders.

Resin cements have excellent aesthetic qualities and are essentially insoluble in mouth fluids. Compressive strength is low, but can be increased by the addition of fillers. They have no inherent adhesion to the tooth. Retention is dependent on mechanical locking when the cement flows into irregularities on the surfaces of the substances being cemented.

Resin cement materials are provided as a two-part powder–liquid product. The powder consists largely of poly(methyl methacrylate) to which various fillers (qv) may be added. These include calcium carbonate [471-34-1], silica [7631-86-9], barium carbonate [513-77-9], and calcium tungstate [7790-75-2]. An organic peroxide, eg, benzoyl peroxide, capable of generating free radicals is also present (see INITIATORS; PEROXIDES, ORGANIC).

The liquid is basically a methacrylate monomer having a suitable inhibitor to ensure adequate shelf life. *N,N'*-Dimethyl-*p*-toluidine [99-97-8] is probably the most common polymerization accelerator although *N,N*-bis(2-hydroxyethyl)-*p*-toluidine and or a sulfinate salt, eg, sodium *p*-toluene sulfinate [873-55-2], also may be used.

Fluid, free-radical polymerizable resin-based cements based on the use of cross-linking monomer systems, eg, BIS–GMA and urethane dimethacrylates, having small particulate fillers, yield strong, insoluble materials suitable for cementation applications, eg, crowns, bridges, composites. Bonding is by penetration into etched or roughened surfaces of the substrate. The use of surface active monomers, such as are being used in dentin adhesive systems, eg, 4-methacryl-ethyloxy trimellitic anhydride (4META), have been successfully used in resin cement formulations.

Visible light activated resin compositions containing calcium hydroxide have been made available as cavity lining materials (99) for pulp capping and pulpotomy, and may offer in other applications an alternative to the calcium salicylate type cements.

Dental Plasters

Gypsum is widely distributed naturally as calcium sulfate dihydrate [10101-41-4], CaSO₄·2H₂O. When partially calcined, the hemihydrate, CaSO₄¹/₂H₂O, is formed (see CALCIUM COMPOUNDS; CALCIUM SULFATES). Gypsum has been used in one form since 1756 for making dental casts, and in another form since 1844 for dental impressions (101).

Plaster is the rehydrated calcined gypsum. The American Dental Association classifies five types of dental plaster according to the physical properties: type I, impression plaster; type II, model plaster; type III, dental stone; type IV, high-strength dental stone; and type V, high-strength,

high-expansion dental stone. These different types are the result of various calcining methods.

Although plaster has been a very successful and serviceable material, it is seriously lacking in hardness, edge strength, chip resistance, abrasion resistance, and strength to fulfill many needs of dentistry. Some of these requirements have been partially filled by the development of the type III and type IV plasters. Table 3 lists the compression strength of dental plasters.

Table 3. Compressive Strength of Dental Plasters, Investments

Materials	Compressive strength, MPa ^{a,b}	ISO standard no.
<i>Plasters</i>		
impression	4.0 ^c –8.0 ^d	6873
model	9.0 ^c	6873
dental stone	20 ^c	6873
high-strength dental stone	35 ^c	6873
<i>Investments</i>		
inlay and crown ^e	2.3 (1.7)	7490
partial denture ^e	5.8 (3.3)	7490
phosphate-bonded ^{f,g}	3.0	7490
ethyl silicate-bonded ^f	1.5	proposed
<i>Dies</i>		
gypsum refractory	13.0	proposed
phosphate refractory	13.0	proposed

^a To convert MPA to psi, multiply by 145.

^b Value given is at room temperature. Value in parenthesis is at casting temperature.

^c Value given is minimum value.

^d Value given is minimum value.

^e Low temperature casting investments.

^f High temperature casting investments.

^g Ref. 102.

To form plaster, gypsum is ground and subjected to temperatures of 110–120°C in open kettles to drive off part of the water of crystallization. The crystals thus formed are large and porous and are the type I and type II plasters. These crystals require a 2:1 powder–gauging water ratio for proper consistency. Type I, impression plaster, and type II, model plaster, differ in additives that control working and setting times. When gypsum is formed these crystals form the type III plaster commonly called dental stone and require 28–35 mL of water for 100 g of powder. Type IV, high-strength dental stone is formed by calcining gypsum in a 30° 0 solution of calcium chloride. The crystals resulting from this process are slightly larger and denser than the type II crystals and require even less gauging water (20–22 mL 100 g of powder). Type V dental stone has gypsum that is formed by a process similar to that for the type IV stone, with the use of calcium chloride. Additives, such as fillers and surface tension reducing chemicals, result in higher setting expansion and higher strength. Hence, less gauging water is required, ie, only 18–20 mL 100 g is needed.

Although gypsum products develop a linear expansion during hardening, the true volume of the final dihydrate is ca 7° 0 less (103,104) than the total volume of the hemihydrate plus the water required for the chemical conversion of the hemihydrate to hydrate. In order to produce a pourable consistency, an excess of water is always used. As the gypsum crystals grow in the form of needlelike projections from each center of crystallization, these growing spiny shapes floating in an excess of water push themselves apart from each other. Voids form between the growing crystals which ultimately interlock to make the mass rigid and hard. Setting expansion is a result of this apparent increase in volume owing to the formation of voids.

As a result of the linear expansion, the reduced volume of the dihydrate, and the evaporation of excess water, the percentage of void spaces in plaster is ca 45° 0, in stone 15° 0, and in improved stone 10° 0. Thus, the additional amount of water required for plaster contributes to the volume but not to the strength of the hardened material (105).

Each 100 g of calcined gypsum theoretically requires only 18.6 mL of water to complete the chemical reaction from the hemihydrate to the dihydrate. Any amount of water greater than 18.6 mL 100 g of powder is excess and reduces the strength of the hardened plaster. When a mixture of the hemihydrate and water hardens, linear expansion takes place. This expansion may amount to as much as 0.5° 0 for plaster. Dental stones also expand on setting, but the amount is significantly less than that permitted in plaster, ie, 0.2° 0 for type III, 0.1° 0 for type IV, and 0.3° 0 for type V.

Impression Plasters. Impression plasters are prepared by mixing with water. Types I and II plasters are weaker than dental stone (types III and IV) because of particle morphology and void content. There are two factors that contribute to the weakness of plaster compared to that of dental stone. First, the porosity of the particles makes it necessary to use more water for a mix, and second, the irregular shapes of the particles prevent them from fitting together tightly. Thus, for equally pourable consistencies, less gypsum per unit volume is present in plaster than in dental stone, and the plaster is considerably weaker.

Impression plasters are formulated to produce a thin, fluid slurry when mixed with the proper amount of water. A satisfactory impression plaster should have a setting time of 4 ± 1.5 min; fineness, ie, 98° 0 should pass a number 100 sieve (ca 0.15 mm), and 90° 0 pass a number 200 sieve (ca 0.07 mm); setting expansion at 2 h should be <0.15%; the compressive strength at one hour should be 5.9 ± 2 MPa (855.5 ± 290 psi); and testing consistency as determined by the diameter of the slump in the consistence test should be 90 ± 3 mm.

Impression plasters are manufactured from the finest finishing plasters, selected for color and purity. Setting time accelerators, setting expansion control agents, fillers, flavors, colors, or other special modifying agents may be added, eg, starch, to cause disintegration of the plaster impression when it is boiled.

Impression plasters are used to obtain an impression (or negative) of the hard and soft tissues of the mouth. The plaster slurry is placed in a tray, inserted into the mouth, pressed in place against the area in question, and held still until it hardens.

Model Plasters. Model plaster should have a setting time of approximately 10 minutes. The fineness of the powder should be such that 98° 0 passes a number 100 sieve (ca 0.15 mm) and 90° 0 passes a number 200 sieve (ca 0.07 mm). Setting expansion should be less than 0.30° 0, compressive strength at the end of one h should be a minimum of 8.8 MPa (1276 psi), and the consistency should form a disk during the slump test of 30 ± 2 mm diameter.

Model plasters are manufactured from select finishing plasters with special emphasis on a clean, white color. Setting-time control agents, setting-expansion control additives, fillers, and pigments may be added.

After it is mixed in water, model plaster is poured into an impression (negative) to produce a cast (positive) or oral structure. This plaster cast is produced chiefly for study or record purposes, ie, successive casts allow an orthodontist to follow and demonstrate the results of corrective treatment. Additional uses for model plaster include the making of study casts and denture repair casts; mounting interarch registration assemblies on articulators; flasking dentures during processing; and for a variety of other applications where strength and abrasion resistance are not of prime importance.

Dental Stones. Dental stones, produced from gypsum, are sold as dry powders in sealed containers. They are prepared for use by mixing with water, in proportions recommended by the manufacturer.

Dental stone is generally used at a water–powder volume ratio of about 30 parts water to 100 parts of stone. The mix is not easily poured, but can flow readily under mechanical vibration. The physical property requirements include a setting time of 10 ± 3 min; fineness of powder, where 98% should pass a number 100 sieve (ca 0.15 mm) and 90% pass a number 200 sieve (ca 0.07 mm); linear setting expansion at 2 h of <0.20%; compressive strength at 1 h of 20.6 MPa (2987 psi); and consistency such that the slump test disk is 30 ± 2 mm diameter.

Setting-time control agents, setting-expansion control agents, fillers, colors, or other modifying additions may be added. Calcium sulfate dihydrate and sodium chloride [7647-14-5] are effective accelerators. Potassium sulfate [7778-80-5] and sodium potassium tartrate [304-59-6] accelerate the setting rate and decrease the setting expansion. Borax [1303-96-4], potassium carbonate [584-08-7], sodium carbonate [497-19-8], and sodium citrate [68-04-2] are all effective retarders and reduce the setting expansion.

Type III dental stones are used for casts requiring higher compressive strength and abrasion resistance than casts formed using the type II plaster. These dental casts are used for the processing of denture-base materials.

High Strength Dental Stone. This dental stone is often referred to as improved dental stone. The crystals are of a cubic or rectangular shape and show a reduction in cracking and porosity. These changes in crystals reduce the amount of water required to wet the powder and produce a workable consistency, ie, 20–22 parts water per 100 parts powder.

Mixes of improved dental stone (type IV) using 22 parts of water to 100 parts of powder produce a mass that is not fluid and pourable but can be easily vibrated into place. The physical properties of the improved dental stone include a setting time 10 ± 3 min, fineness of powder such that 98% passes a 100 sieve (ca 0.15 mm) and 90% passes a 200 sieve (ca 0.07 mm); setting expansion at 2 h limited to a max of 0.10%; compressive strength at 1 h of at least 34.3 MPa (4974 psi); and a disk formed in the slump test for consistency of a 30 ± 2 mm diameter.

Type IV dental stones are used to make casts for a single tooth, for crown or inlay work, and for a complete dental arch.

High Strength, High Expansion Dental Stone. This is a die stone intended for the developments in techniques, materials, and approaches to clinical dentistry that require the higher expansion attainable with model plaster and even higher strength than type IV dental stone. The setting time is 12 ± 4 minutes; the compressive strength must be no lower than 48 MPa (6960 psi); and the setting expansion must lie between 0.10 and 0.30%.

Type V dental stone is used for making dies employed in the fabrication of wax patterns for prostheses cast in nonprecious dental alloys. Nonprecious dental alloys generally solidify at higher temperatures than the high noble and precious metal alloys and have more shrinkage as they cool to room temperature. The higher expansion of type V dental stone provides a larger die and, as a result, a larger pattern that compensates for the greater shrinkage of the casting. Type V stone is a possible choice whenever compensation for more shrinkage of a prosthetic or other material, such as a duplicating material or impression material, is needed.

Dental Investments.

Dental investments are comprised of refractory materials capable of withstanding elevated temperatures. They are used as casting investments and model investments. The compressive strengths, according to International Standards, of the investments and die materials are shown in Table 3.

Casting Investments. Casting investments are used to form molds into which molten metal may be cast. The cavity for receiving the metal is formed by the lost wax process. The composition of investments used for alloys cast from low (<1100°C) temperatures are different from those used for alloys cast from higher (>1300°C) temperatures.

A casting investment must provide sufficient expansion to compensate for shrinkage (up to 0.4%) of the wax (or plastic) pattern during its fabrication, and shrinkage (1.2–2.0%) of the cast alloys resulting from solidification and cooling. The higher the solidification temperature of an alloy the greater is its casting shrinkage.

The compensating expansion of a dental investment may be separated into two types, setting and thermal. Setting expansion includes expansion or dimensional change from the curing of the investment, and expansion from excessive water in contact with the investment after its original mixing with recommended water (106,107). This water may be added by either immersion of the mixed investment in a water bath or as an excess of water from a presoaked porous ceramic liner, called a casting ring, that is placed between the mixture and a metal cylinder container. Thermal expansion is caused by the expansion of the set investment during heating through a wax burnout process and by changes of some component(s) of the investment to lower density phases at the elevated temperatures (108).

Investments for Low-Temperature (T < 1100°C) Casting Alloys. Low temperature casting alloys are usually comprised of gold, silver, and copper and are used for inlays, crowns, and removable partial dentures. At the temperature tested, the compressive strengths specified for inlay or partial denture investments are adequate to prevent fracture during handling and in casting. The strength may vary at casting temperature. The strengths of gypsum-bonded investments at high temperatures vary with the additive used to reduce shrinkage during heating (109). The type of binder also influences the effect of the temperature on strength. Table 3 presents the mechanical property limits defined by ISO standards for casting investment for low-temperature dental casting alloys (110).

These investments consist of a binder of calcined gypsum, ie, calcium sulfate hemihydrate [10034-76-1], CaSO₄·¹/₂H₂O; to which is added silica and modifying agents such as finely ground gypsum, soluble sulfates, and chlorides, ie, accelerators of setting; and low solubility salts, ie, retarders of setting (111). The refractory base consists of the silica in some crystalline form; quartz [14808-60-7] and cristobalite are the main constituents. Quartz is found in great abundance in nature, but cristobalite is prepared commercially by calcining selected quartz at 1500°C. Investments containing cristobalite undergo expansion at lower (T < 600°C) temperatures and also, to a greater extent, at higher temperatures than those containing only quartz. Some of the investments used for casting cobalt–chromium–nickel alloys that solidify at a temperature of about 1300°C also have a gypsum binder. In order to reduce oxidation of these alloys the investments contain carbon [14762-74-4] or colloidal copper [7440-50-8] to produce a reducing atmosphere in the mold (112); this is particularly useful if some previously formed metallic parts are embedded in the investment and are carried through the burnout process.

Gypsum decomposes rapidly, especially in the presence of carbon, at temperatures required for casting many of the cobalt-chromium alloys (112). Accordingly, investments for use with type II cobalt-chromium alloys that have solidifying temperatures about 1300°C must have a binder other than gypsum. Organic and inorganic silicates (113,114) as well as some phosphates (115,116) are used. These types of investments are also recommended for use with high fusing gold and other alloys to which porcelain is fused; and for used with alloys offered as substitutes for the gold alloys. Trapped air between the investment and wax pattern may be a problem, even with the use of a wetting agent and vacuum-investing (117). Although the phosphate investments are preferred for high-melting gold alloys, they are suitable for low melting gold alloys.

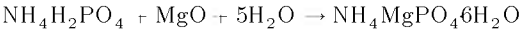
A number of modifiers are employed to prevent most of the shrinkage of gypsum when it is heated above 300°C. Among these are boric acid (118) and sodium and other chlorides (119,120). A review of the development and composition of some investments is given (121).

The minimum total expansion must be 1.5% for casting of inlays and crowns and 1.3% for partial denture gold alloy casting (122).

Investments for High-Temperature Casting Alloys. These investments are used for alloys that require high (>1300°C) casting temperatures. The alloys vary greatly in nobility. Some are rich in gold and/or palladium, and others are predominately cobalt, nickel, and chromium. Some of these alloys are used for crowns and fixed partial dentures with porcelain veneers fused to the metal; others are used for removable partial denture frameworks. The noble-based alloys for fixed partial dentures require casting temperatures in the neighborhood of 1300°C, and phosphate bonded investments are used. The cobalt, nickel, and chromium alloys require casting temperatures of ca 1400–1500°C, and investments that use ethyl silicate or stabilized silica solutions are used.

Phosphate-Bonded Investments. These are available with two components, a powder and a liquid that react when mixed together. The powder consists of refractory materials that control the thermal expansion, eg, silica glass, quartz, cristobalite; setting reaction, eg, magnesium oxide [14762-74-4]; and resistance to reaction with the molten metal, eg, carbon, silica, and other metal oxides. The liquid is a colloidal suspension of silica that provides some green strength and thermal expansion. During the burnout process the silica provides strength through the formation of silicophosphates. The binder is formed

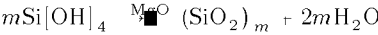
from the reaction of monoammonium phosphate [7722-76-1], $\text{NH}_4\text{H}_2\text{PO}_4$, and calcined magnesia, MgO , in the presence of water.



External measurements indicate that the setting expansion is approximately 0.4° o. A water bath technique adds another 0.6 to 0.8° o expansion when 100° o colloidal silica is used.

The thermal expansion of the investments can vary from 0.8° o using 50° o colloidal silica solution to 1.2° o with 100° o colloidal silica solution. In the green state the modulus of rupture is 0.1–0.5 MPa (14.5–72.5 psi) and ~0.8 MPa (116 psi) fired strength. Firing beyond 900°C results in a decrease in strength due to chemical breakdown of the investment.

Ethyl Silicate-Bonded Investments. These investments are mixtures of powder and liquid. The powder consists of refractory particles of silica glass, cristobalite, and other metal oxides plus magnesium oxide. The liquid is a hydrated silica, tetrasilicic acid [10193-36-9], $\text{Si}[\text{OH}]_4$, that is supplied in a stabilized form; it can be developed by mixing ethyl silicate [78-10-4], denatured ethyl alcohol [64-17-5], and hydrochloric acid [7647-01-0]. The binding of the powder is accomplished by the formation of a silica gel according to the reaction:



Expansion results entirely from heating; there is no setting expansion. In fact, a setting contraction of 0.3–0.4° o occurs. The thermal expansion can be controlled to 1.7–2.1° o up to 1150°C. A heat soak can produce an additional 0.1–0.2° o within one-half of an hour. These investments have very low green strength; models made of them require dipping into a resin that strengthens them for handling. No standard for strength exists yet for these investments. The investments are highly permeable in the as-fired state, and castings of good surface finish, dental accuracy, and fit can be achieved. Fine-grained surface coatings applied to models and patterns produce even smoother surfaces.

Model Investments. Model investments are materials used for noncasting operations in the fabrication of dental prostheses. They differ from casting investments in various ways depending on the prosthetic device being constructed. For low temperature operations, such as soldering, gypsum is used; phosphate-bonded materials are employed for higher solder temperatures or for the fabrication of porcelain veneers.

Soldering (Brazing) Investments. These are either gypsum or phosphate bonded. They are used to hold the metal components of a prosthesis together while they are joined or repaired via a soldering or brazing process (see **SOLDERS AND BRAZING ALLOYS**). The goal is to restrict motion between the parts being joined. Soldering investments differ from casting investments primarily through avoidance of ingredients, such as cristoballite, that cause expansion. The soldering investments contain higher percentages of quartz and other inorganic oxides of metals, such as zirconium [1314-23-4] and titanium [1317-70-0], as refractory components.

Veneering Investments. These are phosphate bonded and contain finely ground quartz, zirconium oxide, and or titanium oxide to produce highly refractory, low expansion dies of fine detail. The dies are formed within impressions taken of teeth that the dentist has prepared in anticipation of covering the front surface with an aesthetic ceramic veneer. Porcelain or ceramic powders are shaped to detail on the dies and these are fired at high (~1000°C) temperatures to produce the veneers. The veneers are then cemented to the front surface of the previously prepared teeth.

Investment Casting Rings and Liners. Casting rings and ring liners are needed prior to investing a wax pattern. Casting rings are cylindrical tubes that mold and retain the casting investment. Ring liners are paper or fabric sheets adapted to the inside of the of the rings to allow the investment to expand within the ring.

Casting Rings. Three types of castings rings commonly used are metal rings, split rings, and disposable rings. Metal rings are the most popular because they can be reused. The rings are available in many sizes and are usually manufactured from stainless steel. Metal rings can be used with most investments, but a liner must be included to allow for expansion. Split rings allow for maximum expansion of investment needed for some alloys. Split rings are made of plastic, stainless steel, or nickel-based alloys and do not need liners. They are available in many sizes. Disposable rings, like split rings, achieve maximum expansion and do not need liners. Most disposable rings are composed of heavy paper and can be cut to different sizes.

Ring Liners. Liners are either cellulose, which readily absorbs water, or ceramic, which does not absorb water or investment liquid (123). Both are available in many sizes. Ceramic liners are made from fibers of alumino-silicate glass derived from kaolin. The principal components in the glass are alumina, 47–65 wt % o, and silica, 35–50 wt % o. Ceramic liners are highly heat resistant to 1300°C. Reports suggest that the fiber from ceramic ring liners may be a possible health risk (124).

Cellulose liner material absorbs water and needs to be wet for at least 30 s prior to investing to prevent liquid from being extracted from the investment. The cellulose liner burns when heated in air and is eliminated completely from the mold at burnout temperatures of 700°C (123). To ensure retention of the mold within the casting ring, a cellulose liner should be kept short of each end. For gypsum investments a liner that is 12 mm shorter than the height of the ring should be used.

Modern liners are all asbestos free because of the health risk associated with asbestos (qv).

Dental Waxes

Waxes and wax composition, ie, pattern, impression, and processing waxes, have been important in the dental profession for many years (125). Most dental waxes find application within the oral cavity during some phase of their use.

Wax cylinders 6.0 mm thick by 10.0 mm in diameter are prepared by a controlled technique, ie, ADA specification no. 4, para. 4.3.1.1. Typical flow curves are shown in Figure 1.

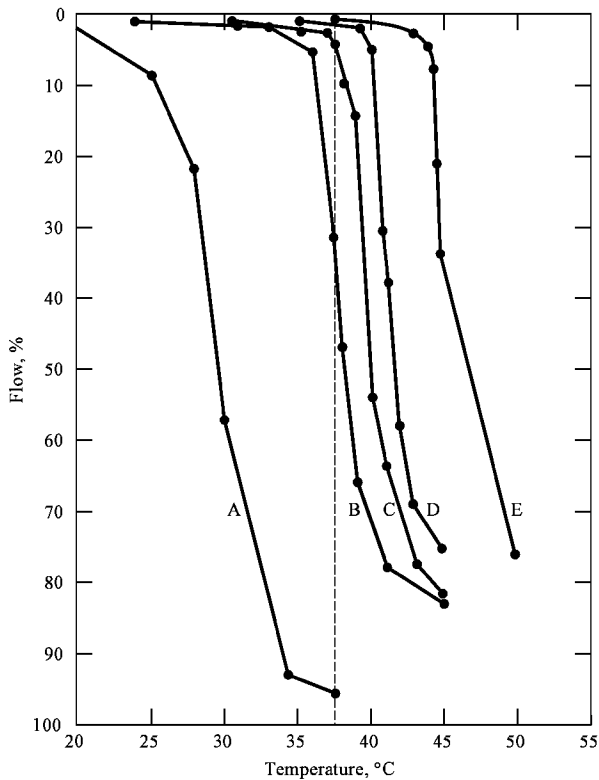


Fig. 1. ADA flow curves of dental waxes. A, impression waxes; B, sheet wax; C, indirect-inlay wax; D, direct-inlay wax; E, base-plate wax. The dashed vertical line represents the temperature (37°C) of the average human mouth.

Flow characteristics at 37°C of a ternary wax mixture of spermaceti [8002-23-1], paraffin [8002-74-2], and ceresin [8001-75-0] are shown in Figure 2 (126). A mixture having composition of 60 wt % paraffin 124, 20 wt % spermaceti, and 20 wt % ceresin is sometimes referred to as the Iowa impression wax (127). It has a melting point of 48.5°C, and a flow of 78.6% at 30°C, 95.4% at 37°C, 96.9% at 40°C, and 97.6% at 45°C.

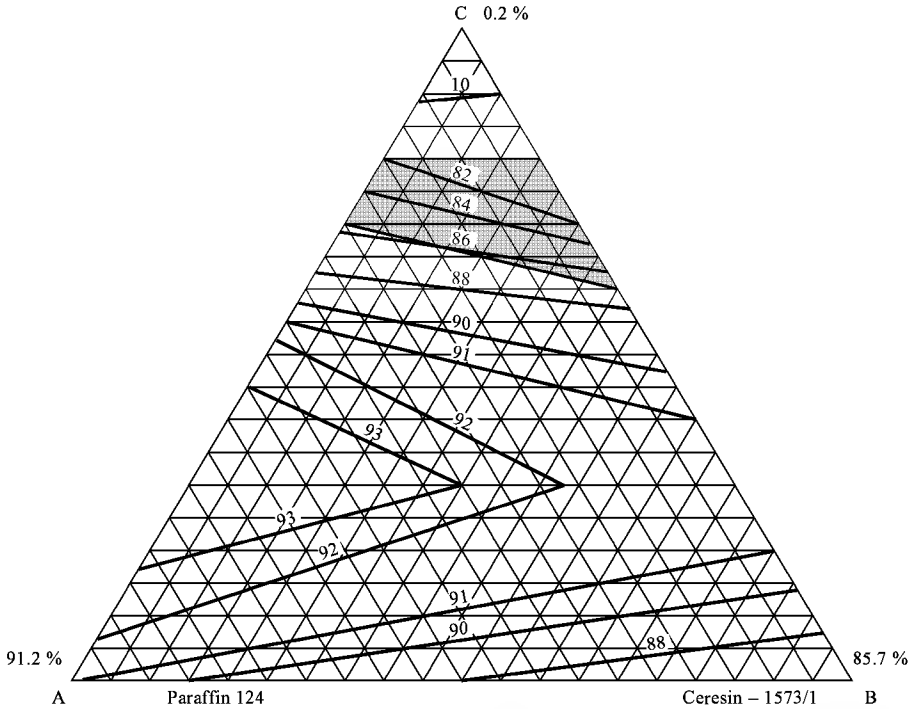


Fig. 2. Flow diagram of ternary wax mixtures of (A) Cornelius paraffin 124, 91.2% flow; (B) Ross ceresin, 1573 1 85.7% flow; and (C) Stevenson spermaceti, 0.2% flow, at 37°C. Numbers in the diagram represent percent flow; the dotted area shows where the waxes are not soluble in each other in the solid state.

Pattern Waxes. The pattern waxes, ie, inlay casting, base-plate, and sheet and shape waxes, are used to construct the prototype or pattern from which a finished dental restoration is produced.

Inlay Casting Waxes. The three types of inlay casting waxes, ie, types A, B, and C, are used to produce wax patterns for the lost wax casting process in the production of cast gold inlays, crowns, and bridges. Some inlay wax is also used to produce patterns for acrylic restorations.

Type A waxes are hard waxes used when an extra hard wax is preferred by the dentist in making patterns in the mouth. Type B waxes are medium waxes used by most dentists to make patterns in the mouth. Type C waxes are soft waxes used for making patterns outside the mouth.

The exact formulations for inlay casting waxes are considered trade secrets, and little has been published on the subject. A binary mixture of 65–75 wt % paraffin wax (60–63°C) and a microcrystalline wax having a melting point >60°C has been suggested (127). This produces a mixture having a solid–solid transition point at about 37°C with little plastic deformation (1–3%) at 37°C and a desirable plasticity at 45°C (73–77%) (128).

Inlay waxes are generally dark blue, green, black, or purple; however, inlay waxes for acrylic work are uncolored and are a natural ivory. The waxes are generally produced in stick form, about 75 mm long and 6.5 mm across; the cross section may be round or hexagonal.

Base-Plate Waxes. Base-plate waxes are used to substitute for, or be used in conjunction with, a base plate to form a pattern for the

production of complete or partial dentures and certain orthodontic appliances to be molded of acrylic resin, modified vinyl resins, or other denture-base polymers.

Base-plate waxes are formulated for specific uses or working conditions into types I, II, and III. Consequently, the flow requirements differ. Type I waxes are soft waxes for building contours and veneers, type II waxes are medium waxes used for pattern production in the mouth in temperate weather, and type III waxes are hard waxes used for production in the mouth in hot weather. At 37°C, type I waxes have a 45–85° flow; at 45°C, type II waxes have a 50–90° flow and type III waxes have a 5–50° flow.

The linear thermal expansion of base-plate waxes should not exceed 0.8° at 25–40°C. It is desirable to invest any waxed-up case as soon after waxing is completed as possible. This minimizes changes in articulation owing to tooth shift, and changes in palatal thickness owing to lifting of the palatal section by wax shrinkage caused by variations in room temperature or by the release of stress.

Base-plate wax compositions are generally regarded as trade secrets. A substantial percentage of paraffin is usually present, probably 50–80 wt %. Beeswax [8012-89-3], carnauba wax [8015-86-9], ceresin, microcrystalline waxes, Acrawax C (Glyco Products Co. Inc.), mastic gum, rosin [8050-09-7], and synthetic resins may make up the balance of the formulation. Base-plate waxes are generally sold in sheet form about 1.3 mm thick, 75 mm wide, and 140 mm long.

Sheet and Shape Waxes. Sheet and shape waxes are used to produce patterns from which complete or partial dentures are cast of gold or base metal alloys. They are used to fabricate the restoration prototype directly upon a refractory investment cast.

The flow of sheet and shape wax is much higher than that of inlay wax and base-plate wax, reflecting increased pliability, ductility, and plasticity. At 35°C these waxes should have a flow of 10° max, and at 38°C a flow of 60° min. At 23 ± 1°C they should not fracture when bent double.

The compositions of sheet and shape waxes are also trade secrets. However, they are blends of various proportions of paraffin, microcrystalline waxes, carnauba wax, ceresin, beeswax, gum dammar, mastic gum, and possibly other resins. Sheet waxes are marketed in square sheets approximately 80 by 90 mm. Various thicknesses are available from 32 gauge (0.5 mm) to 14 gauge (1.63 mm).

Shape wax is similar in composition and properties to the nontacky sheet waxes. It is processed into preformed shapes or definite shape and gauge to facilitate the waxing-up of partial-denture patterns. The shapes and sizes conform to those most needed for lingual bars, palatal bars, clasps, saddle construction, retainers, etc. The available sizes and shapes include rounds from 10 gauge (2.0 mm) to 20 gauge (0.84 mm), half-rounds from 6 gauge (3.36 mm) to 14 gauge (1.63 mm), and half-pear shapes of 6 gauge (3.36 mm). The pieces are usually cut approximately 10 cm long.

Impression Waxes. Impression waxes include those waxes used to obtain a negative cast of the mouth structure (impression waxes), waxes used to establish tooth articulation (bite-registration waxes), and waxes used to detect tooth interference and high spots or improper fit of denture bases (disclosing waxes). They must be plastic and moldable at mouth temperatures, and chill to a still nonplastic mass upon cooling within a few degrees below mouth temperature.

Impression waxes, used for negative casts of mouth structures, are generally considered to be soft, low-melting, plastic waxes used as the last wash on an impression to obtain fine detail of the oral structures. However, complete wax-impression techniques are available that involve the use of hard, high-melting, rigid base waxes, with successively softer wax additions to build a complete impression. The last low-melting wax gives the final detail without thermally destroying the higher-softening waxes of the base impression.

Table 4 lists flow properties of a set of impression waxes the exact compositions of which are trade secrets. The materials that have been identified in the compositions are paraffin, ceresin, vegetable waxes, rosin, mastic gum, and spermaceti.

Table 4. Flow Properties of Impression Waxes

Impression wax	Flow, °	Temperature, °C
grade 1	5	37.5
grade 2	79	37.5
grade 3	28	30.0
grade 4	75	30.0

Bite-Registration Waxes are used to establish the occlusion or horizontal relationship of the lower jaw to the upper jaw when there are opposing teeth present. Bite waxes may have a flow of 84° at 30°C. They are generally compounded from high-flow, low-melting paraffins, microcrystalline waxes, and resins.

Disclosing waxes, or pressure indicating pastes, are used in fitting a dental appliance to establish the location and extent of high spots or pressure areas, as on impressions and complete dentures. They are very soft, salve-like compositions that are painted onto the tissue side of an impression or denture. When the wax-coated denture is seated in the mouth, the soft paste is forced out of the areas showing hard contact between the denture and the mucous membrane. These areas on the denture can be easily marked and material removed from the denture to relieve the premature or hard contact.

Disclosing waxes include soft paraffins, petrolatums, coconut oil, zinc oxide, titanium dioxide, and suitable dyes.

Processing Waxes. The extensive amount of handwork and craftsmanship necessary in the fabrication of most dental restorations and appliances has created a need for several types of wax compositions. These are known as boxing, sticky, utility, or study waxes.

Boxing wax, as the name implies, was developed to box impressions, ie, to construct a retaining ring of wax around an impression to confine the plaster or stone when the cast is poured. These waxes may have melting points of 65.5–71.1°C, and a flow of 60–80° at 37.5°C. Boxing waxes are generally formulated from various microcrystalline waxes to give the desired tackiness and pliability. Various resinous additions may also be made to increase strength and tackiness.

Sticky waxes are used as thermoplastic cements. The broken pieces of a plaster impression are reassembled and held in position with sticky wax. Broken denture bases may be held in proper alignment for repair. Orthodontic appliances may be assembled with a sticky wax prior to investing and soldering. Plaster splints may be sealed to stone models in the production of porcelain or resin facings. Thus, sticky wax is useful in almost any operation where it is desired to position and hold several small pieces in a temporary relationship.

Sticky waxes are generally composed of resins and wax. A high resin content gives viscosity to the melt, a long plastic range, and a brittle fracture when cooled. No modern formulas are available, but the older recipes usually had rosin, beeswax, and gum dammar as the essential constituents.

Utility waxes are waxes of many uses. They are generally useful for sealing, filling, contouring, building up areas, positioning, and many other functional purposes. Their compositions, considered trade secrets, are probably microcrystalline waxes or blends of beeswax, petrolatum, and rosin.

Study waxes, ie, waxes for carving, are useful in the study and modeling of tooth forms and the teaching of anatomical detail. Carving wax compositions include paraffin, ceresin, ozokerite, carnauba wax, montan waxes, and Acrawax C. Fillers and pigments may be added.

Wax Manufacture. Wax compositions for dental usage are usually compounded in simple melting and blending equipment. The melting is done with the least destruction of the wax properties. Direct-fired equipment can be used, but jacketed melting kettles reduce the chance of localized overheating, ie, control of temperatures and heat distribution is desired. Some compositions may require a carbon dioxide or other protective cover to minimize oxidation. Foaming or swelling of the batch in the kettle may occur on melting from thermal expansion of the waxes and resins, raw materials that contain water, volatilization, or low-boiling constituents.

Agitation of the wax melt facilitates heat transfer, reduces localized overheating, and increases the melting rates. Overagitation should be avoided to minimize oxidation of the materials. The melting and processing of waxes should be done as quickly as possible and at the minimum temperature compatible with the process to avoid the loss of volatile constituents, degradation of the materials, and oxidation changes. A cover of inert gas may be used to protect susceptible compositions from excessive oxidation.

Wax compositions are processed into rods, sheets, cakes, and special forms by a variety of processes. Casting, roll forming, and extrusion, by

mechanical or hydraulic-piston-type extruders or by continuous screw extrusion, are operations in practice. Stamping, roll cutting, or molding are also employed.

The applicable specifications for each type of wax are listed below.

Wax type	Specification
Dental inlay casting	ADA no. 4
Base plate	ADA no. 24
Impression	
Bite registration	
Disclosing	
Boxing	Federal U-W-138
Sticky	Federal U-W-149
Utility	Federal U-W-156
Sheet and shape	Federal U-W-140

Dental Alloys and Metals

The chemical and physical properties of some metals and alloys, such as hardness, strength, stiffness, toughness, corrosion resistance, and biocompatibility have provided materials capable of withstanding the most severe demands of restorative dentistry, namely the harsh corrosive environment of the mouth and high stresses on the small cross-sectional areas of teeth. Metallurgical and dental science have given dentistry many excellent alloys for restorative dentistry, prosthetic dentistry, orthodontics, and dental techniques. These alloys include amalgams; base metal alloys such as cobalt–chromium casting alloys (for partial dentures), chromium-containing casting alloys (for crowns and bridges), and titanium alloys (for implants, partial dentures, crowns and bridges); gold and gold alloys, ie, foil, crystal powder, Au–Ca alloy combinations for casting and wrought alloys; platinum and platinum alloys; solders, ie, gold solders, silver solders, and special solders; and electrodeposited metals such as copper and silver. An excellent review on metals and alloys in dentistry is available (129).

Amalgams. The earliest recorded use of amalgam as a dental restorative material was in 1528 (130). Since then amalgam has achieved the status of one of dentistry's main restorative materials. Modern amalgams remain plastic at room and mouth temperatures for several minutes after preparation, allowing them to be compacted into the prepared cavity; they then harden quickly with very small dimensional change. Their proven ability to function adequately and preserve existing dentition is remarkable considering the dental requirements.

Dental amalgam is a novel alloy. Within a few seconds of the start of trituration, the alloy and mercury [7439-97-6] must amalgamate to a smooth plastic mass. Within 3–5 min it must set to a carvable mass and remain so for 15 minutes. Within 2 h it must develop sufficient strength, hardness, and toughness to resist mild biting and chewing forces. It should not tarnish or corrode in the mouth, nor react to produce toxic or soluble salts, and must maintain its color. Most present-day amalgams can be produced to meet a considerable degree of these formidable requirements.

The mercury content in dental amalgam is generally restricted to about 50% or less. Excessive mercury in the final restoration results in a low strength amalgam with poor creep resistance, excessive setting expansion, and poor corrosion resistance. Properly proportioned, mixed, and processed amalgams will have compressive strengths in the range of 210–345 MPa (30,000–50,000 psi) in 24 hours. Tensile strengths are 48–55 MPa (7000–8000 psi). The linear setting expansion from 5 min after the end of trituration to 24 h should fall within 0 ± 20 μm cm. Creep-test cylinders (8 by 4 mm, containing 48% residual mercury after aging for one week at 20°C) shortened from ca 0.9 to 8.0% after being subjected to a static load of 36 MPa (5200 psi) for 4 hours. When the residual mercury content is increased to 53%, the values are 1.52 times as great (131). Attempts to relate either the dimensional change (on setting) or other mechanical properties to clinical behavior have not been very successful, but it has been shown that the creep rate is important in predicting the marginal fracture of amalgam restorations (132).

Composition. The composition of powdered alloys used in preparing dental amalgams usually includes 66.7–75.5% silver; 25.3–27.0% tin; 0.0–6.0% copper; and 0–1.9% zinc. These are commonly referred to as conventional alloys, and the amalgams made from them are conventional amalgams. This composition range was in use for almost a century until two Canadian researchers designed an alloy consisting of a mixture of two powders; one having the customary flakes of the aforementioned composition range and the other, spherical shapes of a silver–copper eutectic, ie, 72% Ag, 28% Cu (133). The ratio of the customary powder to the spherical one was approximately 2:1 by weight, giving an approximate composition of 70% silver, 18% tin, 11% copper, and 1% zinc. Amalgams made with this powder mix have high compressive strengths and low creep, have better clinical performance than conventional amalgams (134), and are sometimes called dispersed amalgams.

Other powders, made from alloys of uniform composition, appeared on the market with approximately the same overall composition but with a high compressive strength of 462 MPa (67,000 psi), high tensile strength of 55 MPa (8000 psi), and low creep (0.2%) when compared to the amalgams made from the mixture of powdered alloys (135). Other alloys having copper content higher than 18% were subsequently developed. The range of composition of some of these alloys is 42.2–70.3% silver, 17.7–30.0% tin, 12.0–27.8% copper, and 0–0.3% zinc. In addition, one of the powdered alloys contained 3.4% indium.

The copper-rich amalgams have performed well in clinical trials in which they were compared with alloys having lower copper content. An improved marginal stability was observed, which may be associated with a longer clinical lifetime (136). These amalgams have also been called non-γ₂ amalgams, where γ₂ refers to the compound Sn_{7/8}Hg comprised of Sn₈Hg [11092-12-9] and Sn₇Hg [11092-11-8].

The complete role of each of the metals in dental amalgams is not fully understood. It is generally agreed that when the conventional powdered alloy is triturated with mercury, the gamma and beta phases of the silver–tin system are attacked by the mercury and that principally Ag₂Hg₃ (γ₁) [12249-78-4] and Sn_{7/8}Hg (γ₂) are formed. These phases form a matrix that binds the remaining unreacted part of the original powder particles together. Usually the less the matrix, the better the values of the pertinent physical properties.

One of the serious defects of the conventional amalgams is the corrosion of the Sn_{7/8}Hg phase, which is normally absent or, if it forms, mostly disappears in amalgams with the increased copper content; hence, the term non-γ₂ amalgams. The phases present in two conventional amalgams used in clinical testing (56,58,62) are

Phase name	Designation	Principal component	CAS Registry Number
Ag–Sn	γ	Ag ₃ Sn	[12041-38-2]
Ag–Hg	γ ₁	Ag ₂₂ SnHg ₂₇	[69484-15-7]
Sn–Hg	γ ₂	Sn ₈ Hg	[11092-12-9]
Cu–Sn	ε	Cu ₃ Sn	[12019-61-2]

In the dispersed-phase amalgam an additional phase of an eutectic of the Ag–Cu system and a reaction ring (zone) of Cu Sn₅ [12019-69-1] around the residual silver alloy particles has been detected. If 10% gold is added to conventional alloys at the expense of the silver content, non-γ₂ amalgams are

formed (137). Other elements, such as manganese and palladium, are also effective in eliminating the tin–mercury compound.

The phases and their proportions present in hardened amalgam are controlled by many factors. The composition of the alloy; the size, shape, and size distribution of the particles; the thermal history of the cast ingot and the comminuted alloy; and the surface treatment of the particles are some of the factors for which the manufacturer is responsible. The tooth cavity preparation and the mixing, compacting, and finishing techniques of the dentist can make the difference between satisfactory and unsatisfactory restorations, even with the best of alloys. A minimal amount of residual mercury and porosity are needed to obtain the most serviceable restorations (138).

Amalgam restorations are prepared by mixing a powdered alloy with mercury to form a plastic moldable mass that is packed or condensed into the prepared cavity. The cavity is designed to provide mechanical retention, maximum marginal mass, support to absorb the functional stresses transmitted through the restoration, and maximum protection to the remaining tooth structure. The restoration reestablishes the normal tooth anatomical form and function.

Amalgams have shown some sensitivity to manipulation variables. It is therefore essential to follow the instructions for proportioning, preparation, condensation, and finishing provided with each individual alloy. The general procedure conforms to the following outline. First, the alloy powder and mercury are proportioned by weight. Current alloy improvements, better trituration equipment, and improved condensation methods have all tended to permit a reduction of the mercury content in the mix and in the restoration. The alloy powder and mercury should be triturated together in mechanical amalgamators with controlling timers as directed. The practice of mulling in the hand following trituration is obsolete and hazardous. There should be very little excess mercury in the mix.

Next, the amalgam must be condensed properly in the prepared, clean, and dry cavity. In general, the maximum effective load tolerated by the patient should be used on a condenser of optimum size, to remove as much mercury from the mass as possible. The amalgam must be carved to the contour of the lost tooth tissue it is replacing.

Final polishing should be delayed for at least 24 h after placing the restoration. Amalgam restorations should not be subjected to biting or chewing forces for at least 2 h (preferably 6 h) after insertion to avoid fracture of the restoration.

A smooth, well-polished amalgam restoration retains its color and appearance much better than a poorly finished one. Roughness promotes uncleanness, discoloration, and corrosion.

Amalgams made with spherical particles may predominate in use over those made with flake-shaped particles because the desirable plasticity is obtained with a lower mercury content, satisfactory compaction is achieved with lower packing pressures, and there is less influence of manipulative variables upon values for appropriate physical properties.

Generally, the newer amalgams having the higher copper content have better values for the properties of interest than the so-called conventional amalgams. The high copper content (non- γ_2) amalgams have less creep, higher compressive strength, lower tensile strength, better marginal integrity in restorations, less discoloration, and are more corrosion resistant.

Manufacturing. In the general procedure manufacturing processes used to produce alloy powders for dental amalgams, metals of the required purity are melted together in their proper proportions under nonoxidizing conditions. Either atomization is used to produce spherical particles or ingots are cast and cooled at a rate that produces some Widmanstätten structure, a geometrical pattern caused by the formation of a new phase along certain crystallographic planes of the parent solid solution. An ingot may be given a homogenization heat treatment at 400°C, and the ingots are reduced to filings by machining on a lathe or a cutter. Iron particles (from the machining operations) are removed magnetically and the filings may be ground to reduce their size. Undesirable particle sizes are screened out to the desired gap-graded sizes. The particles are given a heat treatment at 60–100°C for 1–6 h to pre-age the metal, ie, remove the cold-work effects, and reduce the rate of amalgamation and setting, and the particles may be acid-cleaned to remove any oxide film and other contaminants, washed, and dried.

Gallium-Based Alloys. A gallium-based alloy has been introduced commercially in Japan as a substitute for dental amalgam, but similar alloy types have previously been associated with abnormal cellular reactions and are not much used elsewhere; nickel–gallium alloys have produced carcinomas in rats (139). The corrosion resistance of the gallium alloys is also marginal.

Gold and Gold Alloys. Gold foil, crystal powder, a gold–calcium alloy, and combinations thereof in the noncohesive states are used in dentistry as direct-filling materials; an adsorbed layer of ammonia renders the gold noncohesive for packaging, and heating returns it to a cohesive state for use. Gold alloys are used for cast restorations and prosthetic devices. Wrought gold alloys are used for wire clasps and fabricated orthodontic appliances (see GOLD AND GOLD COMPOUNDS).

Gold and gold alloys serve the needs of dentistry better than any other metals or alloy systems. Gold alloys have a broad range of working characteristics and physical properties, coupled with excellent resistance to tarnish and corrosion in the mouth.

Pure or alloyed, eg, Au–Ca, gold is used as a direct filling material in restorative dentistry as foil (ca 0.635 μm thick), also available as a gold–platinum laminated foil containing up to 40% platinum; and powdered pure gold crystals. Gold, in either of these forms, welds to itself at room temperature when condensed by force. Certain other metals have this characteristic, but the difficulty of achieving and maintaining a sufficient degree of surface purity makes adequate welds difficult to obtain. A gold restoration, foil or other type, of pure or alloyed gold, or both, is relatively weak, soft, and malleable. The cohesive characteristic of the gold foil is partially lost with time as the metal is contaminated by exposure to the atmosphere. The cohesive characteristic is restored, immediately before use, by heating in a clean flame to rid the surface of contaminating substances, including sorbed gases. The gold restoration is produced by malleting or condensing the clean metal into the clean, dry cavity preparation. Hardness of the final restoration exceeds the hardness of cast pure gold, owing to the cold-working. The hardness approaches that of 22 kt gold.

Gold Casting and Wrought Alloys. Gold alloys useful in dentistry may contain gold, silver, platinum, palladium, iridium, indium, copper, nickel, tin, iron, and zinc. Other metals occasionally are found in minor amounts. The effect of each of the constituents is empirical, but some observations have been made.

Gold [7440-57-5], Au, is the principal constituent of gold-colored alloys. It contributes gold color; increases the specific gravity; raises the melting point, if that of the alloy is below that of gold; and increases ductility, malleability, corrosion, and stain resistance. Gold produces heat-treatable compositions with copper, platinum, and zinc. It is useful in amounts of 25–100 wt %.

Copper [7440-50-8], Cu, produces a reddish color and reduces the melting point of the alloy. It produces heat-treatable compositions with gold, platinum, and palladium that result in increased hardness, strength, and generally improved physical properties. The tarnish resistance of the alloy is usually decreased. The gold–copper, Au–Cu, system is the fundamental system of many dental gold alloys. Copper has a useful range of 0–20 wt %.

Silver [7440-22-4], Ag, alters the color of binary gold–silver alloys to green-gold and, as the percentage of silver increases, to a silver-colored alloy. In gold–silver alloys containing copper and platinum, silver improves the color of the alloy. Silver also improves ductility and exerts an influence upon the rate and intensity of response to heat treatment. Excessive silver content may cause alloys to absorb a large volume of oxygen while molten. Upon solidification of the alloy this absorbed gas may be released and cause spitting or gas inclusions in the metal. This problem is seldom evident within the range of 0–20 wt % silver.

Platinum [7440-06-4], Pt, detracts from the gold color, producing an undesirable grayish-red color; increased platinum produces a platinum-colored alloy. Platinum increases strength, proportional limit, and solidification temperatures; reduces grain size; and produces a heat-treatable alloy with gold. It has a useful range of 0–18 wt %.

Palladium [7440-05-3], Pd, produces a large color change, ie, 5–6 wt % may produce a white alloy. It rapidly raises the melting point; increases hardness, tensile strength, and proportional limit of gold–copper–silver alloys; and slightly lowers tarnish resistance. Palladium absorbs large volumes of oxygen and hydrogen when molten, which may cause inclusions in the alloy from oxidized elements, and upon solidification renders a casting porous or defective due to the expulsion of those gases.

Zinc [7440-66-6], Zn, whitens gold alloys, but not rapidly. It lowers the melting point rapidly with increased additions, increases the hardness of gold alloys, increases strength and elastic limit slightly, reduces surface tension, and protects other elements from oxidation. When present in high percentage, zinc embrittles gold alloys.

Iridium [7439-88-5], Ir, and rhodium [7440-16-6], Rh, individually increase corrosion resistance, hardness, and strength of platinum alloys. They can be used to reduce grain size (140).
Indium [7440-74-6], In, is sometimes used as a scavenger of oxygen. It promotes uniform grain size and casting fluidity.
Iron [7439-89-6], Fe, is used to produce age-hardening in Au–Pd–Pt alloys at temperatures involved in ceramo–metallic techniques (141).
Tin [7440-31-5], Sn, is used up to 4.6 wt % to aid in porcelain fused to metal processes and to decrease the melting point of gold solders.
Lead, Pb, and mercury, Hg, are contaminants.

Gold Alloys, Cast Types. Four types of gold alloys have been recognized for cast dental restorations (Table 5). They provide desired material for specific uses. The appropriate specifications for these alloys is ANSI ADA specification no. 5.

Table 5. Mechanical Properties of Cast Gold Alloys^a

Gold cast	Yield strength ^b , MPa ^c	Minimum elongation, %	Vickers hardness, HV
type I	140 max	18	50–90
type II	140–200	18	40–120
type III	200–340	12	120–150
type IV ^d	340 min (500)	10 (2)	150 ^e

^a Values given for annealed specimens, 0.1° offset.
^b Minimum–maximum values.
^c To convert MPa to psi, multiply by 145.
^d Maximum value for the hardened state given in parentheses.
^e Minimum value, 300 HV has been reported.

Type I, soft alloys (20–22-carat golds), are used for inlays of simpler non-stress-bearing types. Type I gold alloys can be burnished, and are not heat-treatable. They are composed essentially of gold–silver–copper with minor modifying additions, eg, zinc.
Type II, medium-hard alloys, are harder, stronger, and have lower elongation than type I alloys. They are used for moderate stress application, eg, three-quarter crowns, abutments, pontics, full crowns, and saddles. The type II gold alloys are difficult to burnish, and can usually be heat-treated.
Type III, hard alloys, are the hardest, strongest, and least ductile of the inlay casting alloys. Their use is indicated for restorations required to resist large forces such as three-quarter crowns, abutments, pontics, supports for appliances, and precision-fitting inlays. These alloys cannot be burnished, and heat treatment improves all their physical properties, except ductility, which is greatly decreased.
Type IV, extra hard (partial denture) alloys, are indicated where high strength, hardness, and stiffness are required. They are harder, stronger, and less ductile than the type III alloys. These partial-denture alloys are used for cast removable partial dentures, precision-cast fixed bridges, certain three-quarter crowns, saddles, bars, arches, and clasps.

Casting gold alloy should be meltable in an air–gas flame, meltable and castable without much change in composition, very fluid when melted with low surface tension, and not absorb gases or oxidize excessively in the molten state. The total shrinkage, from the beginning of solidification to room temperature, should be small. Casting gold alloy should have pleasing, acceptable color, and physical properties, ie, hardness, tensile strength, elastic limit, and fatigue resistance, adequate for the intended use. These physical properties should respond to heat treatment. Finally, casting gold alloy should show resistance to all corrosion or tarnish in the mouth (142) and should be passive (143).
Table 6 gives the range of compositions of the four types of dental gold casting alloys, using both yellow and white golds. Compositional requirements have been removed and replaced by biological testing according to ANSI ADA specification #41. The composition requirements listed are those prior to 1988. In addition to the elements listed, In, Ni, Sn, Ir, and Co have been used as modifying metals.

Table 6. Compositions of Cast Dental Gold Alloys, wt %^a

Type	Characteristic	Au	Ag	Cu	Pd	Pt	Zn
<i>Yellow golds</i>							
I	soft	79–92.5	3–12	2–4.5	0.5 ^b	0.5 ^b	0.5 ^b
II	medium	75–78	12–14.5	7–10	1–4	1	0.5
III	hard	62–78	8–26	8–11	2–4 ^b	3	1
IV	extra hard ^c	60–71.5	4.5–20	11–16	5 ^b	8.5 ^b	1–2
<i>White golds</i>							
III	hard	65–70	7–12	6–10	10–12	4	1–2
IV	extra hard ^{c,d}	60–65	10–15	9–12	6–10	4.8	1–2
IV	extra hard ^{c,e}	23–30	25–30	20–25	15–20	3–7	0.5–1.5

^a Table adapted from Ref. 144.
^b Value given is maximum value.
^c Used for partial dentures.
^d Yellowish white.
^e Very white.

Table 7 gives the composition of gold alloys available for commercial use. The average coefficient of thermal expansion for the first six alloys listed is (14–15) × 10^{–6} °C from room temperature to ca 1000°C; two opaque porcelains used with them have thermal coefficient expansion of 6.45 and 7.88 × 10^{–6} °C from room temperature to 820°C (91). The HV values of these alloys are 109–193, and the tensile strengths are 464–509 MPa (67–74 × 10³ psi). For the last four alloys in Table 7, the HV values are 102–216, and the tensile strengths are 358–662 MPa (52–96 × 10³ psi), depending upon thermal history.

Table 7. Composition of Gold-Based Alloys Used in Ceramo–Metallic Prostheses, wt %

Trade name	Au	Ag	Pt	Pd	In	Other
Amator ^a	82.6	2.4	12.4	0.8		1.8 Zn

Amator 2 ^a	82.0	2.5	11.9	1.8	1.8	
Ceramco ^a	87.7	1.0	6.1	4.6	0.6	
Degudent ^a	84.8	1.3	7.9	4.6	1.25	0.15 Ir
Herador ^a	83.2		15.6		0.9	0.3 Ir
V4 [†]	81.5	2.7	11.7	2.3	1.8	3.0 Ir
Ceramco ^b	87.5	0.9	4.2	6.7		0.3 Fe, 0.4 Sn
Ceramco "O" ^b	87.1	1.3	4.6	6.5		0.3 Fe, 0.2 Sn
Cameo ^b	51.5	12.1		29.5	6.8	
Vivostar ^c	54.2	15.7		25.4		4.6 Sn

^a Ref. 145.
^b Ref. 146.
^c Ref. 147.

Gold Alloys, Wrought Type. Two types of wrought gold alloys were formerly recognized by the ADA specification no. 7 for the fabrication of orthodontic and prosthetic dental appliances, ie, type I, high-precious-metal alloys, and type II, low-precious-metal alloys (gold color). Alloys of this type are seldom used in the United States; they have been replaced by stainless steels and nickel–titanium alloys.

Platinum and Platinum Alloys. Platinum has excellent resistance to strong acids and, at elevated temperatures, to oxidation. Under reducing conditions at high temperatures it must be protected from low-fusing elements or their oxides. Easily reduced metals at high temperatures may form low-fusing alloys with platinum.

Carbon or silicon may produce brittleness and loss of useful properties. Caustic alkalies and many alkaline earth salts or hydroxides may attack platinum at elevated temperatures.

Platinum, a bluish-white metal, is soft, tough, ductile, and malleable. It has a melting point of 1773°C. The coefficient of thermal expansion is 9×10^{-6} °C.

Platinum has many uses in dentistry. Pure platinum foil serves as the matrix in the construction of fused-porcelain restorations. Platinum foil may be laminated with gold foil for cold-welded foil restorations. Platinum wire has found use as retention posts and pins in crown and bridge restorations. Heating elements and thermocouples in high-fusing porcelain furnaces are usually made of platinum or its alloys (see PLATINUM GROUP METALS).

Platinum, as an alloying element, is used in many dental casting golds (Tables 6 and 7) to improve hardness and elastic qualities. Platinum in combination with palladium and iridium has limited use for dental pins and wires.

Palladium and Palladium Alloys. Palladium is not used in the pure state in dentistry. However, it is a useful component of many gold casting alloys, as shown in Tables 6 and 7.

Alloys based on Ag–Pd have been used for a number of years and are available from most gold alloy manufacturers (148). The palladium content is 22–50 wt %; silver content is from 35 to 66 wt %. Minor amounts of Zn, In, or Sn are often present to increase fluidity. Both In and Sn form intermetallic compounds with both Pd and Ag and, therefore, some of the commercial alloys are susceptible to age hardening (149). These alloys are somewhat difficult to fabricate and require meticulous processing. They may also produce a greenish discoloration when they are fused with porcelain veneers. Nevertheless, clinical experience generally has been satisfactory, and cost is the primary criterion for use.

Base-Metal Alloys. Base-metal casting alloys are inferior to gold-based casting alloys in some dental applications but are superior in others. Gold-based alloys are preferred for inlay, crown, and bridge castings because the wax pattern can be reproduced more accurately in the casting. Gold alloys are lower melting and easier to use; restorations can be fabricated conveniently. Up to the 1980s, the noble metal alloys seemed superior for use with dental porcelains. However, base-metal alloys and low-noble alloys have made inroads into dentistry at the expense of the noble metal alloys. Base-metal alloys containing sufficient chromium to make them passive have essentially displaced gold-based alloys in the casting of skeletons for partial dentures (150–153). Most partial dentures are fabricated in commercial dental laboratories from cobalt–chromium, Co–Cr, nickel–chromium, Ni–Cr, or Ni–Cr–Co alloys (Table 8). Some inlays, crowns, or bridges are made of nonprecious alloys. Wrought stainless steels of the 18Cr–8Ni type are used infrequently as denture bases but have supplanted gold alloys in orthodontic appliances for preventing and correcting malocclusion and associated dental and facial disharmonies. Chromium-containing alloys and titanium-based alloys are the only base metal alloys that can be used with dental techniques and have almost no tarnish or corrosion in the mouth.

Table 8. Composition of Base-Metal Alloys for Removable Partial Dentures,^a %

Trade name	Co	Cr	Ni	Mo	Fe	Mn	Si	C	Other
JD	67.0	26.0	0.5	5.0	1.0	1.0			
Vitallium ^b	62.5	30.0		5.0	1.0	0.5	0.5	0.5	
Platinore ^b	60.7	26.7	2.7	5.8	2.6	0.5	0.6	0.3	0.3 W, 0.1 Pt
Niranium N N ^b	64.7	27.5		5.4	0.4	0.5	1.0	0.4	
Ticonium Hard	8.0	16.0	63.0	5.0		3.5			3.0 Al, 1.4 Be
Durallium LG	55.0	27.0	14.0		2.0	1.0		1.0	
Crutanium ^{c,d}		5–15	5–15	0.3	1.0	4–10			
A ^e	43.5	21.6	20.1	7.0	0.25	3.0	0.35	0.05	3.5 Cu, 0.9 Be
C ^e		13.0	68	4.5	2.5				1 Ti, 6 Al, 2 Nb
E ^e	52.0	26.1	14.2	4.0	1.2	0.7	0.58	0.22	

^a These alloys have a minimum yield strength (0.1% offset) of 500 MPa (73,000 psi), a minimum modulus of elasticity of 172 MPa (25,000 psi), and a minimum 1.5% elongation.
^b Ref. 154.
^c Refs. 155 and 156.
^d Balance of alloy is Co.
^e Ref. 157.

Cobalt–Chromium Alloys. Co–Cr and Ni–Cr alloys are used predominately for the casting of removable partial dentures; fixed partial dentures (bridges), crowns, and inlays are also cast. Because of high hardness, corrosion resistance, and wear resistance cobalt-chromium alloys are used for bite adjustments and as serrated inserts in plastic teeth used in full dentures. These alloys are well tolerated by the body and also are used for dental implants and orthopedic implant alloys.

Removable partial dentures are partial dentures that restore function, maintain surrounding dentition in a stable state, and are removable. They require no (or minimal) dental surgery on neighboring teeth, as opposed to the removal of sound, healthy enamel for the insertion of fixed partial dentures

(FPDs). They often have less aesthetically pleasing restoration than FPDs coupled with the nuisance of removing them at night and a psychologically less satisfying lack-of-permanency.

ANSI ADA specification no. 14 provides a requirement for removable partial dentures of a combined minimum of 85% by weight of chromium, cobalt, and nickel or, for alloys failing to meet that minimum, at least 20% chromium. Bio-compatibility is demonstrated by passing the pertinent criteria of ANSI ADA specification no. 41, Recommended Standard Practices for Biological Evaluation of Dental Materials.

Because Co–Cr alloys are difficult to adjust and grind, they find limited use in fixed partial dentures, crowns, or inlays. ANSI ADA specification no. 38 covers alloys used in conjunction with dental porcelains.

The use of Co–Cr alloys is attributed to their light weight, low cost, stiffness, and general passiveness (see CHROMIUM AND CHROMIUM ALLOYS). The high temperatures necessary for proper casting fluidity precludes their use in gypsum-bonded investments; silicate or phosphate-bonded investments had to be developed. Special materials and techniques have evolved to fabricate restorations from the cobalt–chromium alloys, and have essentially restricted the use of these alloys to commercial dental laboratories. Some progress has been made in the development of alloys with lower melting ranges that can be cast in the gypsum-bonded investments using the customary techniques.

Table 9 lists select properties of Co–Cr alloys. It is generally conceded that the casting shrinkage of the cobalt–chromium alloys is greater than that of the gold alloys. The lower density of the base metal alloys provides a weight advantage over the higher-density gold alloys in certain types of bulky restorations. Cobalt–chromium alloys have Knoop hardnesses of 310–415.

Table 9. Comparison of Cast Dental Alloys

Alloys	Linear casting shrinkage, % ^a	Density, g cm ³	Modulus of elasticity, MPa ^b
cast gold	1.2	17–18	95,000
cobalt–chromium	2.3 ^c	8–9	216,000
titanium alloys	1.7–2.0	4.5	117,000

^a Values affected by size and shape of casting. No fixed value for metals.

^b To convert MPa to psi, multiply by 145.

^c Ref. 158.

Nickel–Chromium Alloys. Because gold is so expensive, there was an intensive development in the 1970s and 1980s of alloys based mostly on nickel–chromium with as many as eight modifying elements. Nickel–chromium alloys cannot be cast as easily or as accurately as gold-based alloys, nor can they be fabricated as easily by soldering. A modification of the manufacturer's techniques, ie, using hand instead of mechanical spatulation of the investments, a dry asbestos liner in the casting ring, and hygroscopic expansion, was shown in one study to produce oversize crown castings (0.00–0.45%) (159). Castings made by the manufacturers were consistently undersize by 0.04 to 0.18%.

Ni–Cr alloys primarily contain nickel with ca 12% or more of chromium. They are used mainly for the casting of fixed partial dentures (bridges), crowns, or inlays, either with or without aesthetic porcelain or plastic veneers (Table 10). However, some of them are used for removable partial dentures.

Table 10. Composition of Base-Metal Alloys for Crowns and Bridges^a, %

Trade name	Ni	Cr	Fe	Al	Mo	Si	Mn	Co	Other
Rexillum III	76.0	13.0		3.0	6.0			0.4	1.8 Be
Nobil Ceram ^b	80.75	12.58		3.42	1.53		0.13		0.37 Be, 0.15 Cu
Qualimet ^b	78.51	19.47	0.43	0.21		1.10			
Wiron S ^b	68.96	16.54	0.37	4.15	5.10	0.83	3.05	0.42	
Microbond NP2	66.0	13.0	5.0		7.0	0.75	0.1		
Neydium ^c	79.0	11.0	0.05	1.8	3.6	1.1			3.2 Pd
Ceramalloy ^c	70.0	20.0	0.2		5.6	3.96		0.22	0.02 Ti, 0.09 C
Ticon ^d	70.0	16.0	0.8	3.0	4.1	0.5	3.9	0.9	0.03 C
Microbond NP ^d	76.0	13.8		3.0	5.0	1.1	0.1		

^a ANSI ADA specification no. 38 requires a minimum yield strength (0.2% offset) of 250 MPa (36,000 psi) and a minimum elongation of 2% for ceramo–metallic alloys.

^b Ref. 160. Elements checked but not found: Zn, Cd, In, Bi, Sb, As, Ti, Ga, Ge, Pb, V, W, Nb, and Ta.

^c Ref. 161.

^d Ref. 162.

No data are available on casting shrinkage, but in alloys having the same approximate nickel and chromium contents, pattern-makers usually make allowances for roughly 3% linear casting shrinkage.

Because the melting temperature range of Ni–Cr alloy is 1220–1345°C, it is necessary to heat the investment molds to 800–935°C. The castings should not be pickled in acid because of their high nickel content and should be cleaned by sandblasting. The alloys are generally hard and are difficult to finish and to abrade for clinical adjustment in the mouth.

There is no meaningful correlation between laboratory corrosion tests and clinical performance, but high in vitro corrosion rates on several high nickel dental prosthetic alloys (163), together with the prevalence of allergy to nickel caused by plated jewelry, gives cause for caution.

Titanium-Based Casting and Wrought Alloys. Titanium-based alloys offer an attractive alternative to gold alloys and to the base-metal alloys that contain nickel or chromium. On a volume basis the cost of titanium is roughly comparable to that of the chromium-containing alloys, but the price of titanium tends to be more stable because its ores are abundant and widely distributed (see TITANIUM AND TITANIUM ALLOYS).

Many of the technical problems of fabrication that formerly inhibited the use of titanium alloys in dental castings (164–166) have been effectively solved, and titanium castings may now be obtained for virtually any type of dental appliance at prices that are increasingly competitive. Special induction or electric-arc furnaces are necessary for casting titanium alloys, and this specialized equipment has, until now, been available in only a limited number of commercial dental laboratories. However, the relatively high price of this equipment, attributed to development costs, is expected to decline significantly; this should help to improve the general availability of cast titanium appliances.

The property most frequently cited in connection with the use of Ti dental or medical appliances is titanium's unique biocompatibility. This helps practitioners avoid occasional allergic reactions that occur with nickel or chromium alloys, and removes concerns about the toxic or carcinogenic potential of appliances that contain nickel, chromium, or beryllium. Wrought alloys of titanium are used for orthodontic wires because of their unique elastic

properties, and are used to fabricate most dental implants.

Arc-melted titanium has excellent fluidity and lends itself readily to the creation of thin margins. Sprues must be carefully placed and abundant venting provided, however, to avoid holes and porosity in the casting. The detection of defects by radiography is facilitated by the low density of titanium, and conventional dental x-ray units may be used in many cases.

The relatively high casting temperatures of titanium-based alloys and the chemical reactivity of molten titanium would seem to indicate the use of special investments based on magnesia or zirconia, but more recent experience indicates that certain commercial phosphate-bonded investments produce satisfactory results (167–169). Some progress has been made in producing titanium alloys with lower melting ranges that can be cast in gypsum-bonded investments (170). It now appears that ANSI ADA specification for dental casting alloys can apply to the titanium-based alloys, but there are some areas that still require clarification.

Titanium castings may be soldered and welded using appropriate techniques. The relatively high electrical resistivity of this metal combined with its relatively low thermal conductivity permits heat to be effectively focused where it is needed rather than spread out into adjoining areas. Electrical spot welding is used to join portions of the castings, eg, as in the construction of bridges (see WELDING). The spot welding can be done intraorally without damage to adjoining tissues (171,172) because the tissues do not experience high temperatures. Intraoral spot welding should not be attempted on metals other than titanium because metals that have a higher thermal conductivity than titanium may conduct excessive heat to the tissues.

Properties. The casting shrinkage of titanium alloys is comparable, but not identical, to that of chromium-containing alloys, and greater than that of gold alloys (Table 5). The ability of commercially pure titanium to be cast in relatively thin sections, together with its low density, permits light-weight appliances in bulky restorations.

The Vickers hardness of Ti alloys is strongly dependent on the presence of elements such as oxygen, nitrogen, and carbon that dissolve in hot or molten titanium through contact with residual air molecules or with the surface of refractory molds. A relatively thin (100–250 μm) surface layer is often observed on cast surfaces (173,174) having Vickers hardness values, ie, greater than 350, often associated with brittleness in Ti alloys. The outer portions of this surface layer, which possess the highest hardness values, are ordinarily removed by normal grinding and polishing operations performed on the castings. Below this surface layer, typical Vickers hardness values of 130–180 commonly are observed. Under optimum conditions, castings of commercially pure titanium show yield strengths of 400–500 MPa and ultimate strengths of 490–650 MPa with elongations of 20–30%. Yield strengths and tensile strengths are increased by the presence of dissolved elements.

Titanium alloys should be melted and cast under carefully controlled and optimum conditions. The higher elongation values of titanium castings reflect the superior ductility of titanium. Ductile clasp arms of titanium castings can withstand relatively large bending adjustments without fracture.

The stiffness of pure titanium can be increased slightly by alloying; alloys such as Ti–6Al–4V may be specified for partial dentures requiring additional rigidity. Titanium appliances do not tarnish or corrode in the mouth, have no metallic taste, and are easy to clean because plaque and calculus do not adhere to them. The relatively low thermal conductivity of titanium (relatively close to that of tooth enamel) gives these appliances a seemingly natural feel in the mouth and minimizes thermal sensitivity (175).

Composition. Acceptable composition limits for titanium dental castings have not yet been established, but current practice favors the use of commercially pure or unalloyed titanium. Unalloyed titanium is sold in four grades of purity with grade 1 being the purest and grade 4 the least pure. ASTM specifications govern the maximum allowable content of certain critical elements in each grade. Oxygen is by far the most critical solute element in determining the properties of titanium because of its tendency to produce brittleness. Hardness is a sensitive indicator of the oxygen content of titanium, and limitations on maximum hardness are likely to be part of any future specification.

A need for room temperature ductility motivates the choice of pure titanium rather than a titanium alloy. In appliances requiring greater rigidity or strength, a certain amount of ductility may be sacrificed for extra strength, and a titanium alloy, such as Ti–6Al–4V, may be more suitable.

Oxygen, nitrogen, and carbon promote brittle behavior in all titanium alloys, and it is important to control the cumulative effect of these three elements during casting operations.

Uses. The use of titanium alloys for cast partial dentures offers light weight, low cost, good ductility, adequate stiffness, chemical passivity toward foods and oral fluids, and biocompatibility with the oral tissues.

Casting of Ti alloys for crowns and bridges is done in investment molds that have been allowed to cool almost to room temperatures after firing. The castings are cleaned in electrolytic solutions or in special chemical polishing solutions that impart a bright smooth surface finish to the casting (176,177).

Special low fusing porcelain veneers are applied to pure (unalloyed) titanium dental castings. It is important that firing be done either in a vacuum or inert atmosphere to protect the metal surface from excessive oxidation. The strength of the metal-ceramic bond is apparently adequate although the bonding is thought to involve primarily a mechanical rather than a chemical component.

Stainless Steels (Iron-Based Alloys). The great abundance of iron and the numerous corrosion resistant iron-based alloys, such as stainless steels, have limited uses as restorative dental materials. Dental restorative alloys evolved from gold casting alloys, entrenching the gold casting technology as a preferred method of fabrication of prostheses. The strength and performance requirements for cast alloy prostheses led to the use of alloys, other than stainless steels, from the 1930s–1970s as more desirable, functional replacements for more expensive gold alloys. The iron-based casting alloys of that time did not provide the needed corrosion resistance, especially against pitting and crevice corrosion, casting ease, compatibility with dental porcelains, high temperature oxidation resistance, and surface appearance as did the cobalt– chromium and nickel–chromium alloys.

Today, stainless steels find their primary use in wrought form for temporary applications such as orthodontic wires, brackets, and temporary crowns. The temporary crowns are obtained in preformed sizes shapes and then are trimmed by the dentist with shears to fit over prepared teeth that are awaiting the fabrication of permanent cast crowns.

Orthodontic wires and brackets owe their strength to work hardening during their formation, contrary to the heat treatment process used for wrought gold alloy wires. The 18–8 stainless steels are the most commonly used alloys; their strength, ductility, and elastic modulus are generally higher than the wrought gold alloys. Because the strength achieved by work hardening is lost during the application of heat from a torch, stainless steel orthodontic appliances must be carefully soldered to limit the amount and extent of heat applied along the appliance.

The 18–8 stainless steels contain 17–19% chromium, 7–9% nickel, 0.08–0.2% carbon, 0.75% silicon [7440-21-3] (max), and 0.60% manganese [7439-96-5] (max). Other alloys used for orthodontic crowns are 80% Ni–20% Cr alloy; 40% Co, 20% Cr, 15% Ni, 7% Mo alloy; and nickel–titanium alloy, ie, 55% Ni, 1.5% Co, and balance Ti.

Aluminum. Aluminum [7429-90-5] and aluminum alloys play a limited role in dental applications. Aluminum is used to make lightweight impression trays, articulator facebows, and radiograph alignment holders. Soft aluminum is used as preformed temporary crowns where shaping and bending can be easily done at chairside. An aluminum-containing alloy, consisting of approximately 12% aluminum, 80% copper, 5% iron, 3% nickel, and the remaining magnesium having trace amounts of silica, silver, and chromium, has been used as a low cost crown and bridge casting alloy in some foreign countries, but has not seen extensive utilization in the United States (178).

Electrodeposited Metals. Electrodeposited, electroplated, or electroformed metals are used to produce an accurate metal-clad die, or cast, from a compound polysulfide rubber, or silicone rubber impression (see ELECTROPLATING). Copper and silver are used to give accurate dies and casts with an improved working surface having strength, abrasion resistance, and chip resistance. The acidic copper sulfate plating bath has been more compatible with the compound and silicone rubber impressions than with the polysulfide, eg, Thiokol, rubber impressions. The impression surfaces are rendered conductive by a coating of powdered graphite, powdered metallic copper, or powdered metallic silver. Control of plating bath composition, careful application of the conductive layer to the impression surface, cleanliness, current densities, and other factors are essential to achieving consistent and successful results. Usually 12–18 h are allowed to produce a satisfactory metal deposit of 25–30 μm thickness. The metal shell obtained must be filled or reinforced with some solid base to prevent distortion or collapse. Artificial stone or self-curing resins may be used for this purpose, before the metal shell is removed from the impression.

The metal-plating baths used are acidic copper sulfate and alkaline silver cyanide. Acid contamination in the alkaline silver cyanide bath will release extremely poisonous hydrogen cyanide gas. For this reason, the two plating setups should be isolated from each other. Both plating baths should be well

ventilated and covered when not in use to reduce evaporation losses and contamination.

Copper-plating bath compositions of various types have been used. A typical bath formulation consists of 200 g copper sulfate crystals, 30 mL conc. sulfuric acid, 2 mL phenylsulfonic acid, and 1000 mL distilled water. A pure copper anode may be used; a copper anode containing a trace of phosphorus reduces sludge accumulation in the plating bath.

Silver-plating bath compositions are somewhat variable; a typical composition contains 36 g silver cyanide, 60 g potassium cyanide, 45 g potassium carbonate, and 1000 mL distilled water. A pure silver anode is required (179).

Solders and Fluxes. Dental solders, like all dental alloys, must be biologically tolerated in the oral environment. They are specifically designed or employed for the purpose of fusing two pieces of dental alloy through the use of an intermediate low temperature filler metal.

There are several types of joining processes that involve the use of molten filler metal and are distinguished by the temperatures employed in the joining operations, ie, soldering, welding, and brazing. Alloys used in soldering have the lowest melting points, with liquidus temperatures less than 427°C. Brazing alloys have liquidus temperatures greater than 427°C, and welding alloys have temperatures close to those of the metal pieces to be joined. The soldering process involves the use of solders and fluxes. Solders are materials that fuse and join two dental alloys together. Fluxes coat the surfaces of alloys to be joined in order to produce clean metal surfaces at the temperatures of joining. This enables the molten solder to fuse with those surfaces.

Solders. Modern dental solders are made from mostly corrosion-resistant, nontoxic metals. Minimal quantities of tin and other elements are often added, some of which could produce toxic effects in the unalloyed state. Each solder is used for specific applications (180–188); typical compositions and properties of solders used in dentistry are presented in Table 11. Most of the ingredients of solders are resistant to corrosion, and alloying them with other ingredients renders the alloy safe for use in appliances placed in the oral environment. Silver solders corrode, but are used only for temporary appliances. Available products do not contain cadmium, although cadmium was an ingredient of some silver solders up to ca 1980.

Table 11. Dental Solder Compositions and Applications^a

Solder type	Compositional range, wt %									Temperature ranges	Application
	Au	Ag	Cu	Zn	Pd	Ni	Cr	Co	Sn		
low karat	45	25–35	15–20	0–9					0–1	691–816°C	orthodontic appliances
general-purpose	50–75	10–25	10–22	4–8		0–6			0–2	724–835°C	post-porcelain soldering, joining of noble alloy prostheses, fastening of wire clasps to partial dentures
high karat	80	3–10	8–12	0–1	0–8				0–1	746–871°C	post-porcelain soldering, joining of yellow gold alloy prostheses
nonprecious alloy	30–40				30–40	25–40	0–20			<1150°C	nonprecious porcelain alloy pre-porcelain soldering
						70–85 ^{b,c}		10–15			
						60–90 ^c	3–30	3–30			
						15–25	5–30	40–60			
silver		42–67	15–40	7–20		0–3				ca 607–688°C	stainless steel orthodontic wires, brackets, temporary appliances
porcelain–gold	60–90	10–30	5–10	1–3	5–10	0–1			1–2	<1065°C	joining of noble alloys prior to porcelain application

^a Manufacturers of these include: J. F. Jelenko Co., Armonk, N.Y.; Austenal Dental Products, Chicago; Ivodlar North America, Amherst, N.Y.; J. M. Ney, Bloomfield, Conn.; Johnson & Johnson Dental Products Co., East Windsor, N.J.; Unitek Co., Monrovia, Calif.; and Masel Orthodontics, Bristol, Pa.

^b Also contains 5–10% Fe.

^c Also contains 0–3% Al.

Gold, platinum, palladium, and silver are the principal components of most of the solders used for joining both noble and base metal alloys. Some solders for base metal alloys also contain nickel, chromium, and or cobalt as primary ingredients.

A satisfactory solder must not corrode or tarnish in the mouth fluids, ie, it must be sufficiently noble in composition, and its composition must be such that its solution potential approximates that of the metal upon which it is used. A solder's fusion temperature must be lower (by at least 100°C) than that of the alloy upon which it is employed so that the alloy is not fused during soldering. In the case of wires, the solder's melting temperature must be less than the recrystallization temperature of the metal.

Solders should flow promptly and smoothly over the surfaces of the parts to be joined. This property depends on the surface tension, viscosity, and adhesive properties of the molten solder. Finally, the color of a solder should match that of the metal employed, and its physical properties should be at least as good as those of the metal, in order for the joint not to be a source of weakness (150).

Fluxes. Fluxes, composed mostly of salts or oxides of metals, serve to protect underlying metal from the air. This prevents the formation of surface oxides that impede fusion and the formation of a strong solder joint. Fluxes may also act to selectively leach elements from the surface of the underlying metal. The result is a surface free of obstacles to fusion, and of a composition readily wetted by the solder.

Fluxes also contain agents that facilitate application, such as petroleum jelly or alcohol. These burn or evaporate at elevated temperatures before the soldering temperature is attained, leaving behind a uniform coating of flux. Fluxes become molten before the joining process is reached. The molten flux flows or spreads to form a continuous coating over the surfaces to which they have been applied.

Antifluxes are materials used to coat certain surfaces to prevent solder from flowing to those regions. Antifluxes are not wetted by the solder. They should be removed at the time when the flux is removed.

Uses. Dental solders and fluxes are used to join orthodontic wires, fasten attachments to partial dentures, repair castings units, and join crown and bridge units either before or after the application of porcelain. They may also be used to repair fixed and removable dental appliances.

A solder and its flux are designed to function as a system for joining specific alloys. The first step in any soldering operation is to make certain that the flux and solder used are those recommended for the alloy to be joined. Next, the pieces to be joined must be thoroughly cleaned. Abrasive blasting or grinding with a sandpaper disk are two common techniques. Once the surfaces have been cleaned, they should be aligned using an acceptable procedure such as an occlusal index. The adjoining parts should be contoured to provide only a small gap between the surfaces to be joined. A space of approximately 0.1 mm is desirable. Alignment of the parts on dies or with clamps, followed by stabilization with soldering investment, is appropriate. Flux should then be applied liberally to the surfaces. In cases involving deep or inaccessible joints a thin coating of flux should be applied before the alignment process. Fluxing agents frequently tend to flow beyond the areas intended for joining, causing the solder to fill or obliterate carefully developed contours in cast restorations. The use of an antiflux such as graphite from a soft lead pencil or jeweler's rouge, softened by solvents such as heptane, can contain the flow of the flux and solder. These materials are applied on the surfaces on which no solder is desired.

Heat is applied with a torch or by placing the appliance into a furnace after the proper surfaces have been coated with flux. It is important to heat

the parts slowly and uniformly to prevent warpage or inaccuracy in the alignment of the parts. If a torch is used, a slightly reducing or neutral brush flame is usually employed. It is applied to bulky regions near the surfaces to be joined, but not directly to them. The flame should consist of a sharply defined blue inner cone and a long diffuse outer region. Heat is applied gently to prevent excessive bubbling of the flux while water and other agents, such as petroleum jelly, are being eliminated. As the temperature increases, the flux melts and flows over the entire surface to be joined. The temperature should be raised until the solder flows when it is touched to the surfaces at the thicker section. The solder should flow freely to the thinner areas and wet both surfaces throughout the joint. After the solder has been applied, the flame is removed and the soldered appliance is allowed to cool under ambient conditions well below the solidification temperature. It may then be cooled more rapidly to near room temperature. The cooling rate should be chosen according to a predetermined plan to develop either a softened or hardened appliance. The appliance is then removed from the fixture and thoroughly cleaned. The cleaning methods used may include a combination of abrasive blasting, soaking in hot water or mild acidic solution, wire brushing, or abrasion with a sandpaper disk. Manufacturer's directions should be followed. Care must be taken to remove the flux as most fluxes contain toxic chemicals or compounds that could interfere with subsequent operations such as veneering with porcelain. Flux that remains on bridgework that is yet to receive porcelain can cause bubbles and discoloration in the porcelain. The porcelain-to-metal bond may also be affected. The remaining flux may promote corrosion at the joint.

The same general procedures of heating and cooling are followed in the case of soldering in a furnace, except that a piece of solder can be laid on the junction prior to insertion in the furnace. The solder melts and flows if the upper furnace temperature is properly set. The temperature should be set at ca 14–28°C (25–50°F) above the liquidus temperature of the solder or at the manufacturer's recommended soldering temperature.

Nonconventional Solder Systems. Nonconventional solder systems are developed for use with newer alloys, especially base metal alloys. They are few in number and will probably remain the exception rather than the rule. Some new solder systems consist of metallic particles either pressed to form a rod or suspended in a paste flux. The metallic composition is close to that of the alloy to be joined. If the particles are nonhomogeneous, the solder has particles with melting points lower and higher than that of the alloy. For nonhomogeneous solders, once the flame has been placed on the parts to be joined and the soldering material, it should not be removed until the flow process is completed.

Polymeric Dental Materials

Acrylic Resins. The first synthetic polymer denture material, used throughout much of the 20th century, was based on the discovery of vulcanized rubber in 1839. Other polymers explored for denture and other dental uses have included celluloid, phenolformaldehyde resins, and vinyl chloride copolymers. Polystyrene, polycarbonates, polyurethanes, and acrylic resins have also been used for dental polymers. Because of the unique combination of properties, eg, aesthetics and ease of fabrication, acrylic resins based on methyl methacrylate and its polymer and or copolymers have received the most attention since their introduction in 1937. However, deficiencies include excessive polymerization shrinkage and poor abrasion resistance. Polymers used in dental application should have minimal dimensional changes during and subsequent to polymerization; excellent chemical, physical, and color stability; processability; and biocompatibility and the ability to blend with contiguous tissues.

Denture Bases. Prosthodontics is involved with materials and techniques for the replacement of missing oral and extraoral maxillofacial tissues. A large part of prosthodontics involves replacing the teeth of edentulous patients, eg, making dentures.

Dentures require accurate fit, reasonable chewing efficiency, and lifelike appearance (189). The chewing efficiency of artificial dentures is one-sixth that of natural dentition (190). Acrylic resins are generally used as powder liquid formulations for denture base, bone cement, and related applications. Polymerization is achieved thermally using initiators; photochemically using photoactive chemicals and either uv or visible light irradiation; and at ambient temperatures using initiator activator systems.

Heat-Cured Resins. For optimum comfort and to impede further loss of chewing efficiency, close adaptation of the denture base to contiguous oral tissues is required, which necessitates custom-made appliances. Nearly all dentures are made of acrylic resins. A wax pattern is used to form a custom denture base in which the denture teeth are embedded. A plaster or dental-stone investment split mold of this wax denture base and teeth is prepared. The wax portion is removed and the surface of the resulting mold cavity is painted with a separating medium, usually an aqueous solution of alginate, to aid in the removal of the cured acrylic from the plaster mold.

A compression-molding process unique to dentistry is used to process the denture base. A monomer-polymer dough or slurry consisting, by volume, of two to three parts polymer beads and one part monomer is packed by pressure into the plaster mold that also contains exactly positioned teeth. The monomer consists of methyl methacrylate (MMA) inhibited by an inhibitor, eg, 4-methoxyphenol [150-76-5] or 2,6-di-*tert*-butyl-4-methylphenol [128-37-0] (BHT); small amounts of plasticizers; and other acrylic monomers, eg, 1–5% of a cross-linking monomer such as ethylene dimethacrylate [97-90-5] or allyl methacrylate [96-05-9]. The solid portion contains suspension-polymerized methyl methacrylate [80-62-6] (PMMA) or methyl acrylate [96-33-3], modified by comonomers such as ethyl [97-63-2], butyl methacrylate [97-88-1], or ethyl acrylate [140-88-5] to increase solubility in the monomer; 0.5–1% benzoyl peroxide [94-36-0]; and pigments, dyes, and opacifiers. Particle size and molecule weight distribution of the resin beads control the monomer solubility and the working consistency of the dough mixture. Traces of suspension agent remaining in the beads, eg, poly(acrylic acid) or soluble starch, may prevent wetting of the beads by the monomer. The polymer beads in the dough may contain enough residual benzoyl peroxide to make further addition of this initiator unnecessary. The molecular weight of a typical MMA PMMA heat-cured acrylic dough can be as high as 1,000,000. Crosslinking agents result in network polymers having improved physical properties.

Most acrylic dentures are fabricated from heat-cured bases. The polymerization rate increases with temperature and is directly proportional to the square root of the concentration of the initiator, eg, benzoyl peroxide. A customary thermal curing cycle for the MMA PMMA dough is 90 min at 65°C, followed by heating to 100°C for 60 min or processing for 9 h at 74°C in a water bath. The curing cycle generates enough free radicals to give a denture that shows minimum porosity and is fully cured in the thick portions as well as in the thin palate areas. After cooling the mold assembly in air, the denture is separated from the investing material, trimmed, and polished. Small amounts of red fibrous materials and highly cross-linked, oversized beads of varying translucency often are added to simulate the appearance of oral tissue. Another mode of thermal curing involves the use of microwave techniques (see MICROWAVE TECHNOLOGY). Whereas microwave processing can significantly reduce the curing cycle, care must be exercised to minimize the introduction of porosity in the denture (191–194).

Ambient Temperature Polymerization (Self-Cure). Polymerization at room temperature is feasible using a suitable redox initiator-accelerator system, ie, 0.3–0.8 wt % of a tertiary aromatic amine accelerator such as *N*, *N*-dihydroxyethyl-*p*-toluidine [3077-12-1] is added to the monomer, which is slurried with PMMA containing ca 0.2–2% w/w benzoyl peroxide. Rate and degree of polymerization depend on the type and concentration of initiator and accelerator, the particle size, and the nature of the polymer. Processing by compression molding is very similar to the procedure for heat-cured materials. However, because of the relatively short working time of self-curing resins, the number of dentures that can be cured simultaneously using a freshly mixed monomer-polymer dough is limited.

The temperature rise resulting from the exothermic polymerization reaction depends on the initiator system, the amount of dough used, and the type and concentration of monomer in the mixture. Temperatures within the bulk portions of the denture are higher than those at the surface. The resulting polymer is not porous because no excessive evaporation of monomer occurs. Most of the polymerization takes place within 30 min, but a posture can continue for hours. The denture flask is therefore held under pressure for 2–3 or longer to ensure as complete a cure as possible. Self-curing material usually contains 3–5% free monomer compared with only 0.2–0.5% found in heat-cured denture base. Excessive amounts of monomer in the finished denture adversely affect its properties. Amine-peroxide cured materials are subject to yellowing and have poorer color stability on aging than heat cured or peroxide-sulfinate cured products; stabilizers and uv absorbers can ameliorate this color instability (see UV STABILIZERS) (195). The main advantage of self-curing resins is their high dimensional accuracy, which results from the lower curing temperature; this alleviates the stresses associated with heat cure. The difference in the thermal expansion of the denture resin and the plaster mold may result in dimensional changes when the mold is subjected to wide temperature ranges during processing. Self-cured dentures have better dimensional accuracy (196,197), but heat-cured dentures are also clinically acceptable.

Fluid or Pour-Type Resins. Fluid or pour-type resins are modified acrylic systems that can be cured chemically. A fine-particle-size polymer powder consisting mostly of high molecular weight material is preferred to prevent a rapid increase in viscosity during mixing and pouring. Polymerization occurs in flexible

agar, alginate, or soft gypsum molds. The fluid resin is mixed and poured into the mold through sprues and cured in a pressurized flask at ambient temperatures for 30 minutes. This technique simplifies the laboratory procedure and reduces costs (198). However, extensive shrinkage of the resin tends to pull the posterior teeth from the resilient mold and out of the occlusal pattern; flow of the fluid resin over the necks of the anterior teeth is incomplete and bonding to crosslinked plastic teeth is inadequate (199). The technique can be improved by treating the teeth with a 50:50 mixture of dichloromethane [75-09-2] and methyl methacrylate (200). With proper precaution and careful attention to detail, clinically acceptable dentures can be obtained with this technique.

Injection-Molding Resins. In this technique, a special flask is equipped with an injection chamber and sprues to connect the chamber with the mold cavity (201). The mold space is filled by injecting the softened resin under pressure. Upon cooling, the thermoplastic resin solidifies in the mold. The mold is filled automatically, provided the correct pressure is maintained. The more elaborate and costly injection-molding equipment is suitable for large-scale dental laboratories. A variety of thermoplastic materials, such as poly(methyl methacrylate), vinyl–acrylic copolymers, polystyrene, polycarbonates, and polysulfones, are processed by this method (see POLYMERS CONTAINING SULFUR, POLYSULFONE RESINS; STYRENE PLASTICS; VINYL POLYMERS). Because of inferior wetting, the strength of the denture-base bond to synthetic teeth is less than that obtained with a heat-cured dough system.

Light-Cured Resins. A newer technique employs light curing of a premixed polymer dough (202,203). The molds selected for the arch-form teeth are based, in part, on the arch sizes of natural human dentition. The teeth and denture base are brought together in a specially designed instrument and processed directly on the cast for a few minutes by high intensity visible light in a special curing unit. The denture-base monomer system uses a nonvolatile multifunctional acrylic resin without MMA that forms an interpenetrating polymer network with the polymeric powder component of the dough. The uncured resin includes a photo-initiator system, such as a camphorquinone–tertiary amine system. Dual cure systems also have been developed eliminating the laborious and time-consuming steps associated with heat-cured molding technique. The properties of the cured material are similar to material cured at ambient temperatures or by heat. The technique is cost-effective and reduces the number of patient visits. It is especially useful in the construction of partial dentures and in repair and rebasing procedures.

Properties of Denture-Base Materials. Physical properties of acrylic denture-base materials are given in Table 12 (204). Mechanical properties of denture bases can vary considerably, and depend on composition, mode of polymerization, and degree of interaction with the oral environment.

Table 12. Physical Properties of Denture-Base Resins

Property	Poly(methyl methacrylate)	Vinyl–acrylics
tensile strength, MPa ^a	48–62	51
compressive strength, MPa ^a	75	61–75
elongation, %	1–2	7–10
elastic modulus, MPa ^a	3.8 × 10 ³	2.3 × 10 ³
impact strength, N·m	1050	3150
transverse strength, MPa ^a	41–55	41–55
flexural strength, MPa ^a	83–117	69–110
Knoop hardness	16–22	14–20
thermal coefficient expansion, °C ^{−1}	81 × 10 ^{−6}	71 × 10 ^{−6}
heat-distortion temp, °C	160–195	130–170
polymerization shrinkage, %	6	6
24-h water sorption	0.3–0.4	0.07–0.04

^aTo convert MPa to psi, multiply by 145.

Self-curing resins have lower strength and stiffness, but the same elastic modulus as heat-cured materials. Polymerization shrinkage of the MMA–PMMA and similar systems is ca 6%. Clinically, a linear shrinkage of ca 0.5% across the posterior part occurs upon normal denture processing. The water sorption of the acrylic material in the mouth partially compensates for this shrinkage. The net linear shrinkage of 0.3–0.4% is clinically insignificant because the tissue on which the denture rests adjusts to such changes (201). Some dental resins contain copolymers of hydroxyethyl methacrylate. Such resins impart increased wettability, but high water sorption lowers the dimensional stability. The thermal expansion of denture base can be reduced by fillers. The resulting dentures have higher impact strength, but reduced polishability. They tend to stain and foul because food particles, debris, and bacteria can penetrate into surface irregularities and voids.

Radiopaque materials are used to determine the location of aspirated dentures and fragments (205,206). Opacifying additives include barium sulfate, barium fluoride, barium or bismuth glasses, and brominated organic monomers and polymers. The incorporation of these additives into the resin base or tooth can adversely affect physical properties. Radiopaque materials meeting the requirement for ANSI–ADA specifications for denture-base polymer have been described (207).

High impact strength, increased hardness, lower thermal expansion, and high fatigue strength are also important properties required of denture-base materials. To address these deficiencies, alternatives to the traditional PMMA dentures have been sought. These include the use of other base polymers and reinforced designed denture systems.

Vulcanized rubber, phenol–formaldehyde, and cellulose nitrate resins preceded the use of acrylic polymers in prosthetics. Epoxy resins were found unsatisfactory clinically because of handling and curing problems and their dimensional and color instability. Vinyl–acrylics, polystyrene, polycarbonates, and polysulfones can be injection-molded to yield dentures with outstanding toughness, high fatigue strength, and low water sorption. Polycarbonates excel in impact strength, but processing requires high temperature and the use of porcelain teeth, which leads to crazing in the polymer around the necks of the teeth because of differences in thermal contraction rates.

Acrylics processed by compression-molding techniques give dentures as satisfactory and dimensionally stable as the estimated 5% of the dentures made with special resins and elaborate processing equipment (208). Promising fiber-reinforced dentures have been reported (209–212).

Requirements for heat- and cold-curing denture-base resins prepared from powder–liquids, gels, preopolymerized blanks, and fluid resins composed of acrylic, vinyl, and polystyrene polymers are given in ANSI–ADA specification no. 12 for denture-base polymers.

Special-Purpose Resins, Repair Resins. Fractured acrylic dentures can be repaired with materials similar in composition to cold-cured denture resins. These materials generally cure more rapidly because of the relative simple manipulations involved. The process is quick and there is little dimensional change, but the strength of the repaired denture may be only half that of the original appliance (213). Test methods and requirements of these materials are given in ANSI–ADA specification no. 13 for denture cold-curing repair resins.

Denture Reliners. A denture can be readapted to the changing contours of soft tissue by relining it with rigid or resilient materials. A model of the soft mouth tissues can be used or a relining resin placed directly on the denture. The latter technique has gained considerable clinical acceptance. Some of these materials generate enough heat during polymerization to injure oral tissues. The resins usually contain plasticizers or low molecular weight polymers to facilitate seating of the denture. Even the hard, cold-curing liners that fulfill the requirements of the respective specification are considered only temporary expedients.

Resilient Liners. Resilient liners reduce the impact of the hard denture bases on soft oral tissues. They are designed to absorb some of the energy produced by masticatory forces that would otherwise be transmitted through the denture to the soft basal tissue. The liners should adhere to but not impair the denture base. Other critical properties include total recovery from deformation, retention of mechanical properties, good wettability, minimal absorption of

fluids, nonsupport of bacterial or fungal growth, and ease of cleaning. At present, no material fulfills all of these requirements.

Resilient liner materials are generally supplied as powders, liquids, or ready-to-use sheets. The commercial materials currently available are plasticized acrylics or silicones (see SILICON COMPOUNDS, SILICONES). The acrylics contain higher molecular weight acrylate, methacrylate polymers, or MMA copolymers with ethyl, *n*-propyl, or *n*-butyl methacrylate, and a liquid that may contain ethanol, ethyl acetate, an aromatic plasticizer, and a monomer. The glass temperature of the cured material should be below mouth temperature. Soft acrylics adhere well to dentures, have poor elasticity, and harden upon aging owing to loss of plasticizer (214–217). Reliners with hydrophilic groups are usually based on hydroxyethyl methacrylate and its polymer (215). The cured resin softens and swells when placed in water, which may cause undesirable changes in the contours of tissue-bearing surfaces.

Silicone reliners are supplied as either a one-component system that cures in the presence of moisture or heat, or a two-component system containing base and catalyst. Both types adhere poorly to denture base and cannot be polished satisfactorily. Some silicones support propagation of bacteria such as *Candida albicans*. Acrylic-based siloxane monomers and resins have been proposed for overcoming these deficiencies (211).

Soft reliners can weaken the strength of the heat-cured resin, because they reduce the thickness of the denture base and allow the diffusion of the monomer or solvent from the reliner into the base. Relined dentures stain readily and are difficult to clean. A polyphosphazine fluoroelastomer has also been developed in an attempt to overcome the deficiencies of available liners (218).

Tissue Conditioners. Tissue conditioners are gels designed to alleviate the discomfort from soft-tissue injury, eg, extractions. Under a load, they exhibit viscous flow, forming a soft cushion between the hard denture and the oral tissues. The polymer in tissue conditioners is often the same as that used for resilient liners. The liquid is a plasticizer containing an alcohol of low volatility (219,220).

The alcohol swells the poly(ethyl methacrylate) beads, rapidly promoting diffusion of the plasticizer into the polymer. As a result of the polymer-chain entanglement, a gel is formed. The conditioner is applied to the denture and provides a cushioning effect; alcohol and plasticizer are slowly leached out, and the material becomes rigid. To ensure resiliency, the conditioner must be replaced after a few days. Some materials exhibit high flow over a short period compared with others with low initial flow; the latter remain active longer.

Crown and Bridge Resins. These materials, based on methyl methacrylate, higher molecular weight methacrylates, and the epimine resin system, are used as interim tooth coverage during fabrication of permanent prostheses. They are used to maintain the correct biting relationship, stop teeth drifting, and protect the prepared tooth against pulpal irritation and fracture. The epimine resin system requires cationic polymerization rather than free radical polymerization common to acrylic materials (211,221).

Resins are also used for permanent tooth-colored veneers on fixed prostheses, ie, crown and bridges. Compositions for this application include acrylics, vinyl–acrylics, and dimethacrylates, as well as silica- or quartz-microfilled composites. The resins are placed on the metallic substrates of the prostheses and cured by heat or light. These resins are inexpensive, easy to fabricate, and can be matched to the color of tooth structure. Acrylic facings do not chemically adhere to the metals and are retained only by curing the resin into mechanical undercuts designed into the metal substrate. They have relatively low mechanical strength and color stability, and poor abrasion and strain resistance; they also deform more under the stress of mastication than porcelain veneers or facings.

Plastic Teeth. Excessive wear, crazing, and blanching of the early acrylic teeth, first introduced in the 1930s, has been overcome by better methods of molding and improved crosslinking. Usually more cross-linking agent is employed in acrylic teeth than in denture-base resin. Residual emulsifier and stabilizers employed in the suspension polymerization of the resin are completely removed. Plastic teeth are manufactured by an injection- or transfer-molding process. To fabricate cross-linked anterior teeth, the labial portion of the front veneer of the tooth is made first and partially cross-linked. Then the mold is packed with a dough of a monomer–polymer mixture containing less cross-linking agent. The assembly is cured to give the tooth a solvent-resistant facing. The backside of the tooth, which is only lightly cross-linked, is receptive to the monomer of the denture-base gel, and a strong union results. Both particle size and molecular weight of the polymer, as well as residual resin initiator content, must be carefully controlled; the rheology of the monomer–polymer mix and cure parameters affect the molding process.

In addition to acrylics, vinyl resins, polycarbonate, and polysulfone have been suggested for teeth molding. Newer compositions contain very finely dispersed spheres such as pyrogenic silica as reinforcing fillers, ie, microfillers; a urethane dimethacrylate resin (222); nonfilled highly crosslinked copolymers with interpenetrating networks; and a layered tooth with an exterior containing a BIS–GMA-type resin and silica filler. Pigments impart a natural appearance. The fluorescence of human teeth under uv light is simulated by the incorporation of additives, and a more natural appearance is obtained by painting on stains and striations.

Acrylic teeth have a compressive strength of ca 76 MPa (11020 psi) and a Knoop hardness of 18–20 KHN. These values are lower than those of alloys used for dentures and those of human enamel or dentin. Acrylic teeth have a more natural appearance and are more fracture resistant than porcelain teeth because of their higher impact strength. The low modulus of elasticity reduces the clicking often experienced by porcelain denture wearers. A chemical bond between tooth and resin base is obtained during processing by heat curing. Acrylic teeth have less resistance to cold flow and higher water sorption than porcelain teeth. Bonding of plastic teeth to self-curing resins is poorer than to heat-curing denture base, but can be improved by treatment with a 50:50 mixture of dichloromethane and methyl methacrylate (200). The deficiencies of plastic teeth are their relatively high wear rate and ultimate loss of occlusal relationship. Plastic teeth exhibit less wear than the supporting structure of the denture prosthesis.

Mouth Protectors. The widespread use of protective mouth guards in contact sports has greatly reduced the incidence of orofacial injuries (223,224). Guards are produced from ready-made stock or are custom-fabricated. Natural rubber, poly(vinyl chloride), poly(vinyl acetate-*co*-ethylene), polyurethane, and silicone elastomers are the materials of choice. Custom guards are fabricated from poly(vinyl acetate-*co*-ethylene) blanks, soft acrylic dough, liquid rubber latex, polyurethane, and laminated thermoplastic (225,226). Ready-made protectors are often poorly fitting, leading to discomfort, gagging, and interference with speech. Properly fabricated mouth protectors are dimensionally stable and acceptable to the wearer.

Maxillofacial Prosthetic Materials. Extraoral or external maxillofacial prosthetics (EMFP) is the science of using polymeric biomaterials for the restoration of missing and/or defective facial tissues. The synthesis of materials that can be easily fabricated into lifelike facial devices has long challenged researchers. Ideally, maxillofacial polymeric materials should mimic as closely as possible the mechanical, physical, chemical, and aesthetic properties of the natural tissues they replace. Mechanically, these materials should combine toughness, strength, and durability with pliability, resiliency, and softness. These materials also are required to exhibit an elastic modulus and compliance that matches that of facial tissues over a relatively wide range of ambient temperatures, eg, 20 to 45°C. Other important mechanical properties that have been difficult to achieve are tear and abrasion resistance and adequate adhesion to contiguous tissues via surgical adhesives. Maxillofacial polymers should be biocompatible, dimensionally stable, physically and chemically inert, color stable, translucent, and solvent and stain resistant, eg, have low water uptake and be relatively unaffected by contact with food, oils, dust, detergents, and other cleansing agents. They also should blend with contiguous tissues in appearance, be able to be colored and textured, exhibit low thermal conductivity, and be easily fabricated and repaired with minimal contraction on processing.

Maxillofacial polymers include the chlorinated polyethylenes, polyetherurethanes, polysiloxanes (see ELASTOMERS), and conventional acrylic polymers. These are all deficient in a number of critical performance and processing characteristics. It is generally agreed that there is a need for improved maxillofacial polymers that can be conveniently fabricated into a variety of prostheses (218,227,228).

Other Uses. Patterns for gold-inlay castings can be prepared from acrylic plastics. The finished casting is not superior to casting produced from a wax pattern, however, and the technique has not aroused much interest. Epoxy die materials are used for the fabrication of cast prostheses in some commercial dental laboratories. Cold-curing plastics are used in making contoured impression trays. These resins contain substantial quantities of fillers that increase the rigidity of the materials after setting. Occlusal night guards are made from poly(methyl methacrylate) or another thermoplastic to protect teeth against excessive grinding. Orthodontics makes extensive use of plastics in retainers, splints, temporary space maintainers, and bite plates (229).

Elastomer Impression Materials. Dentistry requires impression materials that are easily handled and accurately register or reproduce the dimensions, surface details, and interrelationship of hard and soft oral tissues. Flexible, elastomeric materials are especially needed to register intraoral tooth structures that have undercuts. The flexibility of these elastomers allows their facile removal from undercut areas while their elasticity restores them to their original shape and size.

An elastic impression material must be easily and quickly prepared; set quickly to an elastic mass in the mouth; not be harmful or cause discomfort to the oral tissues; and flow to all areas without the need of excessive force. It also must copy detail accurately; possess sufficient strength, toughness, and elasticity to resist permanent deformation when removed from the mouth; not adversely affect the set properties of the cast material; be capable of being

manipulated within a temperature range tolerated by the oral tissues; and have sufficient viscosity to remain in a tray.

Impression materials based on natural polymers, eg, reversible hydrocolloids (agar), irreversible hydrocolloids (alginates), combinations of agar alginate, and, to a lesser extent, zinc oxide–eugenol cements, are still employed. However, the growing use of elastic impression materials by the dental profession has spurred a continuous search for better elastomers. The agar- and alginate-based products fulfill a basic need, but owing to their dimensional instability with any loss or gain of water, improved systems have been sought. The primary emphasis has been in the development of nonaqueous elastomeric impression materials.

Aqueous Impression Materials.

Reversible Hydrocolloids (Agar). The agar-based impression materials are thermally reversible, aqueous gels (230,231), that become viscous fluids in boiling water and set to an elastic gel when cooled below 35°C. The popularity of agar-based impression materials has diminished with the introduction of elastic impression materials such as alginate-based, polysulfide, silicone, and polyether impression materials, but agar [9002-18-0] materials are still used in substantial quantities.

Impressions of inlay and crown preparations, and all gingival areas, are best obtained by filling the preparation or gingival area with impression material injected from a hypodermic syringe. This eliminates trapping air in the corners and recesses and gives a more faithful reproduction without nodules or other imperfections.

The agar-based impression materials are used extensively for duplicating casts. Frequently, it is desired to retain the original model for reference and do the actual work on a duplicate cast. Partial-denture fabrication requires that the original stone cast be duplicated in an investment. For duplicating, the agar-based impression material is usually diluted with water, boiled, cooled to the desired temperature, and carefully poured over the model to be duplicated.

The characteristics that make agar good for impression compositions include unusually good elastic behavior and hysteresis between its liquefying and gelation temperatures. Agar sols liquefy at 70–100°C. On cooling, they form a solid gel at 30–50°C. The delay in gelling brings the temperature within a range tolerated in the mouth. Another useful characteristic is the dimensional stability of the material during gelation, cooling, and storage as long as the water content remains constant.

Agar-based impression materials must have a compressive strength of at least 0.2 MPa (29 psi). They should have a strain in compression of 4–20° in stresses of 9.8–98 kPa (1.4–14.2 psi) per specification test method, and should not have a permanent deformation exceeding 3° after 12° strain is applied for 30 seconds.

Agar, a sulfuric acid ester of a galactan complex, is the basic ingredient for agar-based impression. About 6–12° agar is generally used with a 75–85° water content. Fillers may include zinc oxide and clays. Small percentages of boron compounds increase the viscosity, strength, toughness, and resiliency of the composition. Borax, calcium metaborate [13701-64-9], and organic borate compounds have been used; waxes or fatty acids plus emulsifying agents may be present; plaster-accelerating agents are frequently added and may include potassium sulfate, magnesium sulfate [7487-88-9], and zinc sulfate [7733-02-0]. Such additions can be kept to a minimum if the cold-water-soluble salts and organic contaminants are washed from the agar before compounding. Agents for pH control and fibrous reinforcements have been included in formulations. Agar impression materials must be formulated to remain stable in storage. They should be free of salts or additives that crystallize and salt out, degrade the complex agar molecule, or induce syneresis in the agar gel. Any syneresis destroys the dimensional accuracy of the composition.

Alginate-Based Irreversible Hydrocolloids. Alginate impression materials are supplied as a dry powder. When the correct proportions of the powder and water are mixed, a viscous but slightly fluid mass is formed. The gel formation is based upon the conversion of soluble salts, usually potassium or sodium salts of alginic acid [9005-32-7], to insoluble salts of alginic acid, usually calcium alginate [9005-35-0]. Improvements in available alginate products, convenience of use, and a cost advantage have contributed to their increased growth and popularity.

The technique of preparing an alginate impression material requires only the mixing of measured amounts of the powder and water. The paste is placed in a suitable prepared tray and seated over the area of which an impression is desired. Within 1–4 min the material sets to a strong, tough elastic gel. The impression is unseated by a quick, deft movement to minimize tearing or distortion of the impression. The precautions required to stabilize the water content of the alginate impression materials are the same as those described for the agar impression material. The casts should be poured immediately. Most modern alginates do not require treatment in a fixing solution before pouring the cast.

Alginate impression materials are chemically reactive mixtures. All factors that influence reaction rates are, therefore, important in the use of these materials, ie, correct proportioning; temperature of the water, powder, and mixing equipment; and spatulation rate and duration.

Alginate impression materials must have a compressive strength of at least 0.34 MPa (49 psi) 8 min after the start of the mix; at least 3.5 min of this time interval should be in storage at 37 ± 1°C. They should have a strain in compression of 4–20° between stresses of 9.8–98 kPa (1.42–14.2 psi) per specification method and they should not have a permanent deformation exceeding 3° after a 12° strain is applied for 30 s.

Alginate impression materials are usually the potassium or sodium salts of alginic acid. These commercially available materials have various degrees of polymerization, controlled pH ranges, and particle sizes. Alginic acid is derived from specific varieties of kelp, a marine plant. It is a high molecular weight linear polymer of anhydro-β-D-mannuronic acid [1986-14-7].

The cold-water-soluble alginates, ie, ammonium, sodium, or potassium, must be precipitated as insoluble alginate salts to be useful as impression materials. The precipitating compounds can be any salt of a divalent-metal ion where rate of availability for precipitation can be controlled. Calcium sulfate [7778-18-9] (232), in its dihydrate form, has been used most frequently. Lead silicate [11120-22-2] (233) and chromic sulfate [10101-53-8] (234) had been used and may still find limited use; however, their use is contraindicated because of their toxicity. Retardation of the precipitation reaction, to give useful working time, is generally accomplished by the introduction of phosphate compounds; disodium phosphate [7558-79-4], trisodium phosphate [7601-54-9], alkali metal polyphosphates, tetrasodium pyrophosphate [7722-88-5], and sodium hexametaphosphate [10124-56-8] are typical retarders. It is desirable to have the precipitation of the insoluble alginate completed as rapidly as possible, after an adequate working time. The addition of accelerating materials promotes a fast snap over. Aluminum, sodium, potassium, zinc, manganese, and magnesium fluorosilicates are helpful. The inclusion of suitable filler gives bulk to the dry powder, and blends and controls the viscosity and body of the mixed material. Calcium carbonate, magnesium oxide, zinc oxide, diatomaceous earth [7631-86-9], or other silica types may be used.

Nonaqueous Impression Materials.

Polysulfide Impression Materials. In 1953 the first nonaqueous, elastic dental impression material based on the room-temperature conversion of a liquid polymer, a polyfunctional mercaptan (polysulfide), to a strong, tough, dimensionally accurate elastomer, was introduced. The conversion of the liquid polymer to an elastic solid has been achieved in most products by lead peroxide [1309-60-0]. Significant improvements in strength, toughness, and especially dimensional stability of the set polysulfide elastomers over the aqueous elastic impression materials made these materials popular.

The materials are available in three basic grades, ie, light-bodied or syringe, regular, and heavy-bodied types. The products are supplied as a two-part paste system, usually packaged in collapsible tubes. The polysulfide-containing base and the setting-agent paste, usually lead peroxide, are mixed together in approximately equal amounts to give a homogeneous, streak-free mass, which is transferred to a tray of adhesive-coated metal or of custom-made acrylic resin. The filled tray is then placed over the area of interest and held motionless until the rubber has set, usually within 5–10 min.

The polysulfide impression materials can be formulated to have a wide range of physical and chemical characteristics by modifying the base (polysulfide portion), and or the initiator system. Further changes may be obtained by varying the proportion of the base to the catalyst in the final mix. Characteristics varied by these mechanisms include viscosity control from thin fluid mixes to heavy thixotropic mixes, setting-time control, and control of the set-rubber hardness from a Shore A Durometer scale of 20 to 60. Variations in strength, toughness, and elasticity can also be achieved.

The shrinkage on conversion from the fluid state to the elastic solid state has been reported in the range of 0.03° in 24 h for one composition. Generally the shrinkage in 24 h is 0.15° for most products. The strain in compression, 9.8–98 kPa (1.42–14.21 psi), varies from 7 to 15°. The permanent deformation, after 12° strain is applied for 30 s, varies from 4 to 6.5°. The better polysulfide compositions, if properly packaged, remain workable and useful for many years.

The liquid polysulfide polymers, as manufactured, are slightly acidic. In this condition they are very stable. In an alkaline, oxidizing medium they

polymerize rapidly. Heat, moisture, and an alkaline environment accelerate the reaction. Stearic acid [57-11-4], oleic acid [112-80-1], lead stearate [7428-48-0], and aluminum stearate [637-12-7] are active retarders. The polymerization kinetics and rheological behavior of the polysulfides have been studied (235–237).

The polysulfide rubbers are capable of being electroplated by a special technique. Either copper or silver can be used, but the alkaline silver cyanide plating system gives more reliable results.

The polysulfide base material contains 50–80% of the polyfunctional mercaptan, which is a clear, amber, syrupy liquid polymer with a viscosity at 25°C of 35,000 Pas (cP), an average mol wt of 4000, a pH range of 6–8, and a mild, characteristic mercaptan odor. Fillers are added to extend, reinforce, harden, and color the base. They may include silica, calcium sulfate, zinc oxide, zinc sulfide [1314-98-3], alumina, titanium dioxide [13463-67-7], and calcium carbonate. The high shear strength of the liquid polymer makes the compositions difficult to mix. The addition of limited amounts of diluents improves the mix without reducing the set-rubber characteristics unduly, eg, dibutyl phthalate [84-74-2], tricresyl phosphate [1330-78-5], and tributyl citrate [77-94-1].

A number of modifying agents are desirable. Sulfur, in fractional percentages, is a powerful activator. Stearic acid and oleic acid may be used as retarding agents. Additions of acid must be carefully selected because strong acid attacks the methylenedioxy linkages and degrades the polymer. Buffering agents for pH control and perfumes for odor-masking also are used frequently.

The second catalyst paste of the two-paste product is a curing agent. A wide variety of materials convert the liquid polysulfide polymers to elastomeric products. Alkalies, sulfur, metallic oxides, metallic peroxides, organic peroxides, and many metal–organic salts, ie, paint driers, are all potential curing agents.

Many curing systems bring about a liquid-to-solid conversion of the polysulfide polymers. Most curing agents produce an initial solid mass characterized by a high degree of plasticity and poor elasticity. The development of elastic properties is so slow that the materials are not suitable for the accurate reproduction of undercut areas found in oral structures.

Only two types of systems have found application in dentistry. Lead peroxide is the curing agent most frequently used for the polysulfide polymers that serve as dental impression materials. Lead peroxide converts the liquid polymer to an elastic solid within a time short enough for oral applications.

The curing paste is usually dark brown owing to the high concentration (20–78%) of lead peroxide. A nonlead peroxide curing system offers curing pastes in yellow and red that produce lighter shades of set-rubber colors. A suitable vehicle may be selected from the materials listed as diluents for the base. Suitable fillers also may constitute part of the curing paste formula. Buffering agents, thixotropic additives, and other modifiers may be added.

Condensation Silicones. Odor, color, and stickiness of the polysulfide rubbers have deterred universal acceptance. The development of room-temperature-vulcanizing (RTV) silicone rubbers, or siloxanes, has made another acceptable elastomeric system available to dentistry. Materials have a shelf-life of two years or longer, and the setting reaction is completely free from the gas evolution that plagued the earlier versions of these materials. The silicone impression materials have not produced the variety of products that are found with the polysulfide rubbers. Some product individuality has been evident in the initiator systems, but the differences from product to product are less than for the polysulfides.

The silicone bases have a mild, pleasant odor, and are generally white, although some manufacturers add a pink pigment. The materials are nontacky and can be wiped away from instruments, hands, etc, at any stage of the mix or set. The silicones offer a selection of setting rates and mix viscosities. Regardless of base viscosity, the set rubbers are about the same hardness, ie, a Shore A Durometer value of about 55. The linear shrinkage upon curing exceeds that of the mercaptan rubbers and is in the range of 0.2% in 24 hours. The strain in compression, 9.8–98 kPa (1.42–14.21 psi), is 8.8–34%. The permanent deformation, after 12% strain is applied for 30 s, is 1.45%.

Both the silicone base and the catalyst compositions are moisture sensitive and subject to deterioration when exposed to the atmosphere.

Condensation silicone materials are based on hydroxyl-terminated polydimethylsiloxane and are made in the form of two paste or paste–liquid catalyst systems.

Polydimethylsiloxane is of moderately high molecular weight. The silicone is a viscous liquid. Colloidal silica or micronized metal oxides, 5–10 µm particle size, are added to prepare a paste that is mixed with a tetraalkyl silicate containing 50% ethoxy groups, eg, commonly tetraethyl orthosilicate, and 1–2% of an organic tin activator. This type of catalyst has a limited shelf life because of oxidation.

In the presence of the organic silicate, the heavy-metal salts trigger the chain extension and cross-linking reactions that lead to silicone rubber and volatile ethanol as a byproduct. Useful metal soaps include stannous octanoate [1912-83-0], zinc octanoate [557-09-5], dibutyltin dilaurate [77-58-7], and dibutyltin diacetate [1067-33-0]. The reactivity of the different salts varies considerably. Stannous octanoate effects a cure in 0.5–2 min; zinc octanoate may require 24–96 h; the dibutyltin dilaurate, 10–20 min. Heat and moisture accelerate the curing rate, but to a lesser degree than in the case of the polysulfide rubbers.

Addition Silicones. Perhaps the most important development in the area of elastic impression materials has been the addition siloxane (or silicone) system. Several reviews have been published on the materials (238,239).

Addition siloxanes are the kind most frequently used in the dental clinic and are supplied as a two-paste system. One paste contains a low molecular weight silicone with terminal vinyl groups and reinforcing filler, and the other consists of a hydrogen-terminated siloxane oligomer, filler, and chloroplatinic acid catalyst; mixing gives a crosslinked elastomer. Hydroxyl-containing silicones evolve hydrogen in the setting reaction. Some formulations contain palladium to absorb this gas. Addition silicones have the best elastic properties and lowest dimensional change on setting of all elastomeric impression materials.

The silicone impression materials are very compatible with gypsum products, give casts having excellent hard surfaces, and can be electroplated with either copper or silver. However, the acidic copper sulfate bath gives more acceptable results.

The technique of taking a silicone rubber impression is similar to that employed for the polysulfide–rubber impressions. The materials are available in four viscosities; low, medium, high, and very high. The products are supplied as two-part products. The silicone bases are viscous fluids, usually packaged in collapsible metal tubes, or in jars for the heavy-bodied. The catalysts are either liquid or paste products; the liquids are packaged in plastic bottles or collapsible metal tubes. Further reduction in technique sensitivity has been achieved by the use of dual-injectable, automatic mixing syringes. The impression procedures are similar to those given for the polysulfide impression materials.

Polyether Impression Materials. A polyether-base polymer elastomeric impression material was introduced in 1964 (240,241). This material is related to the epimine resin and cured by the reaction between aziridine rings, which are at the ends of branched polyether molecules. The main chain is probably an ethylene oxide–tetrahydrofuran copolymer [27637-03-2]. Setting is by cross-linking brought about by an aromatic sulfonate ester catalyst (242,243) that produces cross-linking by cationic polymerization via the imine end groups. These polyethers are supplied as two pastes. Just as with the rubber and silicone light-body materials, one mix is used in the syringe, and the remainder in the tray. Setting time for these materials is relatively short, ie, less than 2.5 min.

To improve the rheological properties and extend the very short working time, a simple polyester is included as thinner. Mixing is easy, and dimensional change in air is less than 0.1% over several hours. Elastic recovery and reproduction of detail are excellent. The elastomeric cyclic imine impression materials have a higher modulus of elasticity than the condensation silicone or polysulfide rubbers, and are more difficult to remove from the mouth. The materials have relatively low tear strength and an equilibrium water sorption of 14%; thus, polyether impression materials tear readily. Because of their poor dimensional stability in water, they should be stored in a dry environment.

A new type of impression material that has a polyether backbone with urethane–acrylic end groups and that cures free-radically by visible light irradiation, has been developed (244).

Restoratives. Polymer resins were introduced as tooth restorative materials in the early 1940s. These materials can be classified as unfilled tooth-restorative resins, composite or filled restorative resins, and pit and fissure sealants.

Unfilled Tooth Restorative Resins. Unfilled resins were some of the first polymer materials introduced to repair defects in anterior teeth where aesthetics were of concern. They have been completely replaced by the filled composite resins that have overcome the problems of poor color stability, low physical strength, high volume shrinkage, high thermal expansion, and low abrasion resistance commonly associated with unfilled resins.

Composite or Filled Tooth Restorative Resins. Improvements in the properties of resin-based restoratives, brought about by the addition of silane-treated inorganic fillers to unfilled resins, has made these the primary anterior restorative material used today.

They are used by direct placement in tooth cavities or by custom fabricating using composites on a gypsum model of the cavity and then cementing the restoration into the tooth cavity.

The addition–reaction product of bisphenol A [80-05-07] and glycidyl methacrylate [106-91-2] is a compromise between epoxy and methacrylate resins (245). This BSI–GMA resin polymerizes through a free-radical induced covalent bonding of methacrylate rather than the epoxide reaction of epoxy resins (246). Mineral fillers coated with a silane coupling agent, which bond the powdered inorganic fillers chemically to the resin matrix, are incorporated into BSI–GMA monomer diluted with other methacrylate monomers to make it less viscous (245). A second monomer commonly used to make composites is urethane dimethacrylate [69766-88-7].

Composite resins can be cured using a variety of methods. Intraoral curing can be done by chemical means, where amine–peroxide initiators are blended in the material to start the free-radical reaction. Visible light in the blue (470–490 nm) spectrum is used to intraorally cure systems containing amine–quinone initiators (247). Ultraviolet systems were used in some early materials but are no longer available (248). Laboratory curing of indirect restorations can be done by the above methods as well as the additional application of heat and pressure (249,250).

The composites are dispensed as paste systems in one or two parts. Light initiated composites are systems containing one paste that can be directly placed into a cavity and cured by proper light exposure. Generally, complete curing can be accomplished in under one minute, depending on the size and complexity of the restoration. Chemically cured systems are supplied as two-paste systems where the two parts are hand mixed before placing in cavity. Polymerization occurs in 3 to 5 minutes from mixing. There also are dual-cure systems, supplied as two pastes that are mixed, placed, and can be cured with light in under one minute; if not exposed to a curing light, these will chemically cure in 5–9 minutes. Ingredient list of typical composites are available (251).

The combination of a high molecular weight monomer, BIS–GMA, with a bonded filler reduces the total polymerization shrinkage of composites to 1.5–3 vol % (252). In addition, the coefficients of thermal expansion have been reduced and strength and durability enhanced by the addition of the inorganic fillers (252). Fillers can be composed of particulate quartz and silica glass, colloidal silica, and or ceramic particles that are chemically bonded to the methacrylate resin using the silane coupling agent [3-(methacryloxy)propyl]trimethoxysilane [2530-85-0]. Filler particles made from barium-, strontium-, zirconium-, or zinc-containing glass can be used to increase the x-ray opacity of the composite. Filler particles range in size from less than one micrometer to approximately 10 micrometers in diameter. Composites formed only from submicrometer-sized particles are generally classified as microfills. Composites formed from a blend of two filler types, eg, colloidal (submicrometer-size) silica plus small-sized radiopaque macrofillers, with particle sizes ranging from 1–10 micrometers, are classified as hybrids. Composites formulated from only macrofillers larger than about 3 micrometers are generally classified as macrofills. A new class of composite, characterized by a continuous distribution of particles of one composition, has been introduced. Filler loading levels as high as 60% by weight for microfills to near 88% by weight for hybrids and macrofills are feasible (253). ADA specification no. 27 is for direct filling resins.

Pit and Fissure Sealants. The BIS–GMA or urethane dimethacrylate portion of the composite restorative material has been further diluted with methyl methacrylate monomer or other low viscosity monomers and used to seal developmental pits and fissures on natural tooth enamel. These naturally occurring pits and fissures are very prone to decay soon after the teeth erupt. Sealing can be accomplished by properly acid-etching the enamel within and surrounding these pits and fissures and flowing the low viscosity resin onto the etched surface and polymerizing it. The resin flows into the porosities created by etching and polymerizes by light or chemical initiators to form an impervious barrier that has proven to be clinically effective in reducing the decay rate on these vulnerable surfaces (254–258). Pit and fissure resins are covered by ADA specification no. 39 for pit and fissure sealants.

Adhesives. The fundamentals of adhesion have an important role in modern restorative dentistry (see ADHESIVES). The retention of restorative materials to tooth structure, in addition to holding the restoration to the teeth, seals the interface between the tooth and restorative material. Maintenance of this seal reduces pulpal irritation and the potential for recurrent decay. The application of high performance adhesives reduces the need for mechanically retentive cavity designs, thereby minimizing removal of healthy tooth structure. Adhesion also plays an important role in holding prosthetic materials to one another. Polymer adhesive materials can be classified as composite or filled resin cements, porcelain or ceramic coupling agents, metal coupling agents, or enamel and dentin adhesives.

Composite or Filled Resin Cements. Composite resin cements are similar to restorative composites. They are used to adhesively retain crowns, bridges, orthodontic appliances, and other intraoral prostheses to hard tooth structure. The resin cements are similar to the microfil classification of composites, generally with a slightly lower filler level, ie, approximately 50% by weight. The lower filler loading allows these materials to be used in thin films as cements and luting agents; a resin cement has been introduced with a much higher filler content which purportedly retains cementlike rheology. Curing can be by light, chemical, or dual initiating systems. These cements are usually used in conjunction with a dentin and enamel adhesive, metal coupler, and or porcelain or ceramic coupling agent to effectively laminate the prosthesis to the tooth structure.

Porcelain or Ceramic Coupling Agents. Dental porcelains or ceramics are commonly employed as aesthetic restorative materials. Restorations fabricated from these materials can be adhesively retained on tooth structure or metallic prostheses using the proper combination of couplers and adhesives. Repair of fractured or damaged porcelain restorations can be made using composite restorative materials in conjunction with these coupling agents. Adhesion to the porcelain or ceramic surface is achieved through an organofunctional silane, 3-(methacryloxy)propyltrimethoxysilane [2530-85-0] (259,260). The silane is a low concentration solvent solution that is supplied as either a single-component prehydrolyzed system or a two component unhydrolyzed system that must be mixed to achieve hydrolysis prior to application. Porcelain or ceramic is usually etched with hydrofluoric acid and painted with the silane coupling agent. The silane interacts with the porcelain or ceramic surface and chemically copolymerizes with the composite cement used to lute it to the tooth or prosthesis (261).

Metal Coupling Agents. Metal prostheses can also be bonded to tooth structure or other restorative materials using composite cements with the proper metal conditioning and coupling agents. Metal conditioning can consist of etching the metal surface chemically or electrolytically, air abrading the surface, or electroplating. Etching is done to achieve a micromechanical surface profile that dental polymers can flow into and mechanically interlock to the surface (262,263). Air abrading with an aluminum oxide abrasive increases surface roughness and area, and provides for a clean oxide formation on exposure to air (264). Electroplating is generally reserved for highly precious alloys that do not readily form oxide layers. Brush plating can be done after air abrading to electrodeposit a thin layer of oxide forming metal, such as tin, on the precious metal surface (265).

Once a metal surface has been conditioned by one of the above methods, a coupling agent composed of a bifunctional acid–methacrylate similar to a dentin adhesive is applied. This coupling material is usually supplied as a solvent solution that is painted over the conditioned metal surface. The acidic functional group of the coupling molecule interacts with the metal oxide surface while the methacrylate functional group of the molecule copolymerizes with the resin cement or restorative material placed over it (266,267).

Enamel and Dentin Adhesives. Tooth enamel and dentin are not homogeneous tissues. Enamel has an extremely highly inorganic content while dentin is relatively low in inorganic material but higher in water and organic structures. The microstructure of the two tissues is also very different; enamel is very dense, homogeneous, and has 96% mineral (inorganic) content, while dentin is porous, heterogeneous and has about 60% mineral content. Simultaneous adhesion to both structures is very difficult but makes up the basis for conservative restoration of diseased tooth structure.

Like metal and porcelain, enamel and dentin must be conditioned prior to application of an adhesive. Conditioning treatments enhance adhesion by cleansing the surface or selectively removing some of the organic or inorganic portions of the dental tissue. Enamel is usually conditioned by etching with some form of acid. Acid etching increases the surface area and roughness, and reduces the surface energy to permit the adhesive resin to wet the enamel and form microscopic tags or projections into the surface irregularities for mechanical retention (268). Acid etchants for enamel are usually supplied as concentrated solutions either gelled or in liquid form. The most common is 37 or 40% phosphoric acid that is placed on the enamel surface for 15 seconds, then rinsed off with water; the surface is dried before application of a resin, eg, BIS–GMA or urethane dimethacrylate, adhesive (269).

Adhesion to dentin also requires the conditioning of the surface. Dentin is most often etched with an acidic agent prior to the application of an adhesive monomer. Acid etching of dentin is only a recently accepted procedure. The acidic conditioner can be supplied as a liquid or gelled form of several acids, ie, phosphoric, nitric, and maleic, or it can be a bifunctional monomer with an acidic pendant group such as a carboxylate or phosphate. The etching procedure removes the surface-altered layer created during cutting or instrumenting, and dissolves the mineral matrix from around the collagen fiber

network to expose a fiber-rich demineralized layer. This fiber-rich layer is then reinfiltrated with an adhesive monomer that polymerizes around the collagen fibers, thereby mechanically retaining the adhesive on the surface (270). The monomer infiltrated fiber layer is called the *hybrid layer*.

Dentin adhesion is dependent upon proper wetting of the collagen-rich surface. Wetting can be promoted through the use of multifunctional coupling agents that adsorb onto the surface of the dentin. Coupling agents are usually monomers with surface active groups such as carboxylic, phosphoric, or hydroxy groups that adsorb to the surface; a portion of the molecule that does not adsorb presents a surface that is more easily wetted by the methacrylate restorative material placed over it. An example of a dentin coupler is *N*-(2-hydroxy-3-methacryloxypropyl)-*N*-phenylglycine [4896-81-5] (NPG–GMA) (271,272). Many surface active molecules are available for dentin adhesion and most are associated with a high degree of hydrophilicity, because of the adsorbing acid group. They also contain a methacrylate polymerizing group that covalently bonds to the subsequently applied restorative resin by a free-radical mechanism. These adhesive monomers are supplied as solvent solutions, or mixtures of monomers that are painted over the conditioned dentin surface. The adhesive film can be polymerized using light, chemical, or dual initiator systems similar to those used in the restorative polymers.

An additional polymer adhesive system is the zinc polycarboxylate cement system that offers good adhesion under moist conditions (273). This cement system contains 90–95° zinc oxide [1314-13-2] and some magnesium oxide powder that is mixed with poly(acrylic acid) [9003-01-4] or an acrylic acid–itaconic acid copolymer. The setting mechanism results from the neutralization of the polycarboxylic acid with metal oxides, as well as the stiffening of the polymeric network by ionic interactions between carboxylate groups with zinc ions and by the reinforcing action of excess metal oxide (274). Poly(acrylic acid) interacts with the Ca²⁺ of tooth structure to form ionized carboxyl groups (275). The adhesion of the cement arises from the etching of the tooth surface by the acid, providing mechanical interlocking, and by the formation of ionic bonds between the calcium and carboxylate ions at the tooth–cement interface. The cement adheres well to tooth structure and many metals, but does not bond to porcelain or dental restorative resins.

A similar material to the polycarboxylate cement is the glass–ionomer system. This material is available as a luting cement, restorative material, or cavity lining material. Restorative and lining ionomers are fast setting while luting cements have slower setting times and smaller filler particles. The system contains a powder of calcium fluoroaluminosilicate glass that is mixed with poly(acrylic acid) or an acrylic acid–itaconic copolymer. The setting mechanism is two phase. The faster reaction is a result of dissolution of the glass by the acid, with calcium from the glass powder forming a calcium polysalt. A slower reaction of aluminum, from the glass, with the acid forms an aluminum polysalt. The set material is a composite of unreacted glass particles bound within a matrix of silica gel and calcium and aluminum polysalt (276–278). Adhesion of the cement to tooth structure is by ionic interaction of the poly(acrylic acid) with the Ca²⁺ of the tooth and etching of the surface by the acid. The cement adheres to tooth structure and many metals, but does not bond well to porcelains. Dental restorative resins can be bonded to glass–ionomer by acid etching the ionomer surface and then using an enamel adhesive (279).

Abrasives

Dental abrasives range in fineness from those that do not damage tooth structure to those that cut tooth enamel. Abrasive particles should be irregular and jagged so that they always present a sharp edge, and should be harder than the material abraded. Another property of an abrasive is its impact strength, ie, if the particle shatters on impact it is ineffective; if it never fractures, the edge becomes dull. Other desirable characteristics include the ability to resist wear and solvation.

It is essential to select the appropriate abrasive for specific purposes. For example, levigated alumina imparts a high polish to metal without removing much metal, but if used on tooth, an excessive amount of tooth tissue will be removed without obtaining the desired polish. The term polishing usually denotes the gradual reduction in the size of surface irregularities or scratches by using successively finer abrasive. The final abrasive should leave a mirror-like high luster.

Dental abrasives can be classified either according to their use or according to the degree of their ability to abrade (see DENTIFRICES). The use classification, adopted for the ADA specification no. 37 for powdered dental abrasive materials, is based on removal of stain from natural teeth or on restorations of all types. Several abrasives are used in dentistry in a variety of grit sizes and shapes.

Aluminum Oxide. Emery [57407-26-8] is a natural oxide of aluminum with various impurities. One of these impurities, iron oxide, also acts as an abrasive. Pure aluminum oxide is made from bauxite [1318-16-7] and has partially replaced emery.

Silicates. Garnet [12178-41-5] is any one of many siliceous combinations with aluminum, cobalt, magnesium, iron or manganese, or both. Garnet for dental use is usually coated on paper with a glue and formed in disks and strips. Pumice is of volcanic origin. Kieselguhr is the siliceous remains of minute aquatic plants known as diatoms. Its common name is diatomaceous earth, and in a coarser form it is used as a filler in many materials. The finer particles are excellent mild abrasive and polishing agents.

Tripoli is a mild abrasive and polishing agent from porous rocks near Tripoli.

Rouge is a fine red powder of ferric oxide [1309-37-1] (Fe₂O₃). It is usually used in the cake form but is also impregnated in paper or cloth known as crocus cloth.

Tin oxide is a pure white powder. Mixed with water, glycerol, or alcohol into a paste, it is used for polishing teeth and metallic restorations.

Chalk is a calcium carbonate prepared by precipitation. It is used in many polishing compounds including dentifrices.

Sand and other forms of quartz are used as a powder in sandblasting. If a gentler abrasive material is wanted, powdered walnut shells are often used.

Carbides and diamond [7782-40-3] are imbedded as small particles in a binder and are used to cut tooth structure. Diamond chips are the hardest and most effective abrasive for tooth enamel.

Zirconium silicate [10101-52-7] is a polishing agent with the unique characteristic of gradually reducing its abrasive capability during use. Its high coarseness when first applied permits it to cut rapidly.

Therapeutic Dental Materials

Fluorides. Most worldwide reductions in dental decay can be ascribed to fluoride incorporation into drinking water, dentifrices, and mouth rinses. Numerous mechanisms have been described by which fluoride exerts a beneficial effect. Fluoride either reacts with tooth enamel to reduce its susceptibility to dissolution in bacterial acids or interferes with the production of acid by bacterial within dental plaque. The multiple modes of action with fluoride may account for its remarkable effectiveness at concentrations far below those necessary with most therapeutic materials. Fluoride release from restorative dental materials follow the same basic pattern. Fluoride is released in an initial short burst after placement of the material, and decreases rapidly to a low level of constant release. The constant low level release has been postulated to provide tooth protection by incorporation into tooth mineral.

Composite Resins. Many composite restorative resins have incorporated fluoride into the filler particles. One commonly used material, yttrium trifluoride [13709-49-4], is incorporated as a radiopaque filler to aid in radiographic diagnosis, and is also responsible for slow release of fluoride from the composites (280). This same effect is achieved with a barium–alumina–fluoro–silicate glass filler in composite filling and lining materials. Sodium fluoride [7681-49-4] has also been used in composites by incorporating it into the resin matrix material where it provides long-term low level release (281–283).

Glass-Ionomers. Glass-ionomers show fluoride release at levels that are usually higher than those found in composite materials. The fluoride is found within the aluminosilicate glass, which is melted with fluoride fluxes and ground to form powder filler. The fluoride is added as calcium fluoride [7789-75-5], aluminum fluoride [15098-87-0], and sodium fluoride [7681-49-4] in a combined proportion of approximately 20° by weight in the final powder (284,285).

Other Fluoride Containing Materials. Many forms of fluoride are used strictly as preventive or therapeutic materials. Varnishes containing sodium fluoride are available to place in a cavity prior to filling (286). These varnishes can also be applied as protective agents on tooth crown and root surfaces. Cavity rinses containing stannous fluoride [7783-47-3] and or acidulated phosphate fluoride are applied topically to a cavity just before filling. The largest group of therapeutic fluorides are the topical formulations used for treatment of both primary and permanent teeth. These materials are topically applied to tooth surfaces as gels or rinses. The current (1992) ADA accepted materials have fluoride present as 1.23° acidulated phosphate-fluoride, 2° sodium fluoride, or 8° stannous fluoride (287). Prescription fluoride supplements are supplied as rinses, drops, and tablets containing sodium fluoride or

acidulated phosphate fluoride in varying doses (288). These are used as both topical treatments and dietary supplements. Fluoride is also available at varying concentrations in many over-the-counter mouth rinses, toothpastes, and chewable vitamins. The most common and widespread form of fluoride is that contained in drinking water supplies. Fluoride is found endogenously in most water supplies at widely varying concentrations. High concentrations of fluoride in water are associated with mottling or staining of enamel, while low levels have been shown to produce no adverse effects (289). Many centralized water systems are currently supplemented to a level of 0.7 to 1.2 ppm fluoride; this has demonstrated a 30% to 40% reduction in dental decay (287–290).

Chlorhexidine Gluconate. Chlorhexidine gluconate [18472-51-0] (1,1'-hexamethylene bis[5-(*p*-chlorophenyl) biguanide] di- D -gluconate) is used as an antimicrobial against both aerobic and anaerobic bacteria in the oral cavity. It is used as a therapeutic supplement in the treatment of gingivitis, periodontal disease, and dental caries. A mouth rinse form is available as a 0.12 wt% aqueous solution (288).

Calcium Phosphate Materials

Hydroxyapatite. The mineral of teeth and bone comprises impure forms of hydroxyapatite (HA) [1306-06-5], $\text{Ca}_{10}(\text{PO}_4)_6(\text{OH})_2$. Because of its excellent biocompatibility, synthetic HA, in ceramic forms, has been used clinically for filling bony defects since the mid-1970s. Dense and porous types of ceramic HA are available. Dense HA ceramics are made by compressing calcium phosphate solids from a very fine powder form, known as the green state, into a pellet which is then subjected to a heat treatment that causes the powder particles to fuse together by means of solid-state diffusion (290). Porous HA ceramics are derived from certain species of coral in the genus *Protos* which have regularly patterned 200 μm diameter pores. Processes have been developed to convert the calcium carbonate [471-34-1], CaCO_3 , in the coral to HA while conserving the macroporous structure (291). Porous ceramic HA permits bone ingrowth if the pores have minimum diameter of 100 μm . Interconnecting pores of 40 to 100 μm are necessary for the development of functioning haversian systems (292). Neither the dense nor the porous HA undergo significant resorption or degradations *in vivo* (293). Several reviews (294–296) provide a thorough survey of the literature concerning the use of both types of ceramic HA for filling periodontal defects or for the augmentation of alveolar ridges that have resorbed due to aging or disease.

Other Ceramic Calcium Phosphate Materials. Other ceramic calcium phosphate materials for repairing bony defect include β -tricalcium phosphate (β -TCP) [7758-87-4], β - $\text{Ca}_3(\text{PO}_4)_2$, and biphasic calcium phosphate (BCP) ceramics which consist of both β -TCP and HA. Unlike ceramic HA, β -TCP resorbs in the tissue (293). The *in vivo* dissolution of BCP ceramic implants was shown (296) to increase with increasing β -TCP : HA ratio in the implants. Both β -TCP and BCP can lead to new bone growth to various extents depending on the applications and the type of materials used (293,296).

Calcium Phosphate Cements. Self-setting calcium phosphate cements have been a subject of considerable interest in recent years. Materials that are totally biocompatible and also harden like a cement at the site of application are highly desirable in a wide range of biomedical applications. Data in the literature (297) show that cementation can occur in mixtures containing a variety of calcium phosphate compounds. The products formed in these systems included a dibasic calcium phosphate known as dicalcium phosphate dihydrate (DCPD) [7789-77-7], $\text{CaHPO}_4 \cdot 2\text{H}_2\text{O}$ (298), octacalcium phosphate (OCP), $\text{Ca}_8\text{H}_2(\text{PO}_4)_6 \cdot 10\text{H}_2\text{O}$ (299), and HA (300).

The cement system that uses either tetracalcium phosphate (TTCP) [1306-01-0], $\text{Ca}_4(\text{PO}_4)_2\text{O}$, and dicalcium phosphate anhydrous (DCPA) [7757-93-9], CaHPO_4 , or TTCP and DCPD (300), as the starting ingredients has been studied quite extensively. This cement forms HA with a stoichiometric Ca-to-P ratio of 1.67; the pH of the cement is at or near the physiological pH (301). These probably contribute to the cement's high biocompatibility and excellent *in vivo* characteristics (302,303). When implanted in bone, the cement resorbs slowly and is replaced by new bone tissue without loss of volume (304). This cement has a compressive strength of over 70 MPa (10150 psi) (301), making it acceptable for repairing bony defects in most nonload-bearing skeletal sites. The cement has been evaluated for a number of dental and medical applications including root canal filling (305,306), alveolar bone implant (307), frontal sinus obliteration (308), and cranioplasty (309). Although calcium phosphate cements have not yet been used clinically other than on an experimental basis, the unique combination of self-hardening property and excellent *in vivo* characteristics suggest usefulness for these materials in a wide range of clinical applications.

Dental Implants

The use of dental implants is increasing due to improved success rates. Success for dental implants is based on immobility, no radiolucency between bone and implant, no bone loss, and no other adverse symptoms such as violations of the mandibular nerve canal. The 1988 National Institutes of Health Consensus Development Conference on Dental Implants (310) provided a review of the history and status on dental implants. Reports indicate that the use of better materials and technology has resulted in the improved implant : bone interface responsible for the higher success rate for dental implants. A number of present day implants have means for mechanical fixation such as grooved and porous surfaces. In 1986, based on these factors, a success rate of 85% after five years and 80% after 10 years was expected (311); this was a 10% increase over a 1979 expectation. Modern dental implants can function for 10 years or more (312). A presentation by an implant manufacturer indicated a success rate of 95% over a five year period for a porous coated titanium alloy implant (313). Titanium was chosen (314) as the material for dental implantation studies involving correlation of implant surface preparation and avoidance of surface contamination with preparation of the bone and the histology of the bone that grew adjacent to the implant.

Regulation. Dental implants are regulated by the Food and Drug Administration. All dental implants fall into the FDA class III which covers devices that are life sustaining, life supporting, or are implanted into the body and have the potential to cause unreasonable risk, illness, or injury. Devices in class III are required to have applications for premarket approval (315). There are 15 to 20 companies that have FDA marketing clearance for specific dental implants, based on substantial equivalency to implants marketed prior to 1976, and approximately one third of these companies are foreign. Marketing clearance is not the same as premarket approval.

Requirements. Requirements for dental implant materials are the same as those for orthopedic uses. The first requirement is that the material used in the implant must be biocompatible and not cause any adverse reaction in the body. The material must be able to withstand the environment of the body, and not degrade and be unable to perform the intended function.

There are a number of designs for dental implants. The patient must have ample bone and the proper design should be chosen for each individual. Another factor that affects the success of dental implants is patient care of the implant and dental hygiene.

The American Society for Testing and Materials (ASTM) F4 Committee on Medical Materials and Devices has developed specifications for chemical composition, mechanical properties, and other factors. Standard test methods also are available from ASTM, 1916 Race Street, Philadelphia. The quality of castings is important for dental implants, and standards to define this would be useful.

Materials. Titanium is the most widely used metallic implant today, and titanium-6aluminum-4vanadium alloy is widely used. Materials used as dental implants throughout history include natural teeth from man, animal teeth, ivory, wood, plastic, vitreous carbon, alumina, and various metals including gold, aluminum, 316L stainless steel, cobalt–chromium–molybdenum alloys, titanium and titanium-6 aluminum-4 vanadium alloy. The cobalt–chromium–molybdenum was used successfully for many years and still is used for casting individual implants and for some of the blade implants. Ceramic materials including hydroxyapatite, high density alumina, and single-crystal sapphire (alpha alumina) also are being studied and : or used as dental implants.

Hydroxyapaite, the mineral constituent of bone, is applied to the surfaces of many dental implants for the purpose of increasing initial bone growth. Some investigators believe that an added benefit is that the hydroxyapatite shields the bone from the metal. However, titanium and its alloy, Ti-6Al-4V, are biocompatible and have anchored successfully as dental implants without the hydroxyapatite coating.

There are a number of questions that must be answered in order to maximize the benefits of hydroxyapatite, eg, how thick the coating should be; how should the coating be applied to the implant; is the chemical composition on the implant the same as before the application; what is the nature of the titanium or titanium alloy hydroxyapatite interface, how strong is it, and by what means is it bonded; and what is the interfacial microstructure?

Surface preparation of the dental implant prior to implantation will have an effect on corrosion behavior, initial metal ion release, and interface tissue response (316). The titanium and titanium alloy dental implants in present use have many forms to assist bone ingrowth attachment including cylinders with holes, screw threaded surfaces, porous surfaces, and other types of roughened surfaces. Methods used to produce porous surfaces include arc plasma

spraying of metal onto the implant and the sintering of spheres to the implant surface.

Types of Dental Implants. Indications for a specific type of implant are based primarily on the amount of bone available to support the implant. Also to be considered is the implant proven most successful. Three types of implants are discussed here.

Endosseous. The implants are anchored in the bone and osseointegrated, and are the most successful implants to date. Endosseous implant types are root form, ie, cylinder, screw, cone; bladeform, ie, plate; and ramus frame. The root form, which can be a hollow cylinder, a screw, or a blade, is the most successful. One technique is to submerge the implant in the alveolar ridge and cover it with mucosal tissue. The implant can be a screw type that has a healing-screw in the top. After healing has occurred, the healing-screw is removed, and the tooth fixture is screwed into the space vacated by removal of the healing-screw. This procedure results in the implant being well covered with bone, and helps avoid both the growth of the epithelium down the sides of the implant and subsequent void space and infection. The modern screw-type implant has a 95% success rate over a period of five years (313).

Subperiosteal. The subperiosteal implants are placed on the residual bony ridge and are not osseointegrated. This implant is most commonly used in the mandible but sometimes is used in the maxilla. Subperiosteal implants have been installed since the 1940s (311) and still have a success rate after five years of only 50 to 60%. A success rate of over 90% for five years and 50% for 15 years also has been quoted (312). Subperiosteal implants are fitted by casting, which is an individual procedure. The casting can be coated with a porous metal coating or other coating and then put in the patient. This may result in an improvement for these implants.

Transosteal and Staple Implants. This implant goes through the bone and is for the purpose of attaching dentures. The mandibular bone staple plate has replaced the transosteal pin. Since 1975 the five-pin staple bone plate has been used predominantly, and in 1985 showed a success rate of 93% after five years (317).

Certain commercial materials and equipment may be identified in the article for adequate definition of subject matter. In no instance does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, or that the material or equipment is necessarily the best available for the purpose.

BIBLIOGRAPHY

"Dental Materials" in *ECT* 1st ed., Vol. 4, pp. 894–928, by F. H. Freeman, Kerr Manufacturing Co.; in *ECT* 2nd ed., Vol. 6, pp. 777–847, by F. H. Freeman, Kerr Manufacturing Co.; in *ECT* 3rd ed., Vol. 7, pp. 461–521, by G. C. Paffenbarger and N. W. Rupp, American Dental Association Health Foundation, Research Unit, National Bureau of Standards.

1. The Standards Coordinator, The American Dental Association, Chicago.
2. F. J. J. Clarke, in J. W. McLean, ed., *Dental Ceramics: Proceedings of the First International Symposium on Ceramics*, Quintessence Publishing Co., Chicago, 1983, p. 441.
3. G. J. Christensen, *J. Prosthet. Dent.* **56**, 1 (1986).
4. D. W. Jones, *Dent. Clin. North Am.* **29**, 621 (1985).
5. D. E. Southan, in J. A. von Fraunhofer, ed., *Scientific Aspects of Dental Materials*, Butterworths, London, 1975, p. 277.
6. W. J. O'Brien, *Dental Materials: Properties and Selection*, Quintessence Publishing Co., Chicago, 1989, p. 397.
7. B. O. Mysen, *Structure and Properties of Silicate Melts*, Elsevier, Amsterdam, 1988, p. 99.
8. J. R. Mackert, in J. D. Preston, ed., *Perspectives in Dental Ceramics: Proceeding of the Fourth International Symposium on Ceramics*, Quintessence Publishing Co., Chicago, 1988, p. 53.
9. **U.S. Pat. 3,052,983** (Sept. 11, 1962), M. Weinstein and A. B. Weinstein (to Permament Manufacturing Co.).
10. W. H. Murrmann, M. Brandestini, and F. Lutz, *J. Am. Dent. Assoc.* **120** (6), 517 (1990).
11. D. G. Grossman, in W. H. Murrmann, ed., *International Symposium on Computer Restorations*, Quintessence Publishing Co., Berlin, 1991, 103.
12. J. R. Kelly and co-workers, in W. H. Murrmann, ed., *International Symposium on Computer Restorations*, Quintessence Publishing Co., Berlin (1991), 253.
13. J. K. Dong and co-workers, *Int. J. Prosth.* **5**(1), 9 (1992).
14. R. B. Sozio and E. J. Riley, *J. Prosthet. Dent.* **49**, 182 (1983).
15. J. W. McLean and T. H. Hughes, *Br. Dent. J.* **119**, 251 (1965).
16. V. Piddock and A. J. E. Qualtrough, *J. Dent.* **18**, 227 (1990).
17. R. R. Seghi, T. Daher, and A. Caputo, *Dent. Mater.* **6**, 181 (1990).
18. D. E. Southan, *Aust. Dent. J.* **15**, 507 (1970).
19. K. J. Anusavice, C. Shen, and R. B. Lee, *J. Dent. Res.* **71**, 1134 (1992).
20. P. W. McMillan, *Glass Ceramics*, 2nd ed., Academic Press, Inc., London, 1979, p. 1.
21. D. G. Grossman, *Dent. Clin. North Am.* **29**, 725 (1985).
22. L. A. George, F. C. Eichmiller, and R. L. Bowen, *Ceram. Bull.* **71**, 1073 (1992).
23. D. G. Grossman, *J. Am. Ceram. Soc.* **55**, 446 (1972).
24. G. C. Paffenbarger, W. T. Sweeney, and A. A. Issacs, *J. Am. Dent. Assoc.* **20**, 1960 (1933).
25. G. E. Servais and L. Cartz, *J. Dent. Res.* **50**, 613 (1971).
26. S. Crisp, M. A. Jennings, and A. D. Wilson, *J. Oral Rehab.* **5**, 139 (1978).
27. A. D. Wilson, *Chem. Soc. Rev.* **7**, 265 (1978).
28. D. C. Smith, R. D. Norman, and M. L. Swartz, in *Restorative Dental Materials; An Overview*, Vol. 1, Quintessence Publishing Co., London, 1985, pp. 35–39.
29. K. DeGroot and co-workers, *J. Dent. Res.* **59**, 1493 (1980).
30. D. C. Smith, *Br. Dent. J.* **124**, 381 (1968).
31. **U.S. Pat. 3,655,605** (Apr. 11, 1972), D. C. Smith (to National Research Development Corp.).
32. D. C. Smith, *J. Can. Dent. Assoc.* **37**, 22 (1971).
33. G. S. Beagrie, J. H. P. Main, and D. C. Smith, *Br. Dent. J.* **132**, 351 (1972).
34. D. R. Beech, *Archs. Oral Biol.* **17**, 907 (1972).
35. D. R. Beech, *Br. Dent. J.* **135**, 442 (1973).
36. D. R. Beech, A. Solomon, and R. Bernier, *Dent. Mater.* **1**, 154 (1985).
37. B. E. Causton, *J. Dent. Res.* **58**, 1383 (1979).
38. B. E. Causton, in *Biocompatibility of Dental Materials*, Vol. 2, CRC Press, Inc. Boca Raton, Fla., 1982, Chapt. 7.
39. M. A. Monharram and co-workers, *J. Mater. Sci.* **21**, 1681 (1980).
40. T. Nasu, *J. Biomed. Mater. Res.* **20**, 347 (1986).
41. K. Miyazaki and co-workers, *J. Dent. Res.* **69**, 367 (1990).
42. Y. Matsuya and co-workers, *J. Dent. Res.* **71**, 141 (1992).
43. A. D. Wilson and B. E. Kent, *Br. Dent. J.* **132**, 133 (1972).
44. A. D. Wilson and H. J. Prosser, in *Biocompatibility of Dental Materials*, Vol. 3, CRC Press, Inc., Boca Raton, Fla., 1982, Chapt. 3.
45. A. W. G. Walls, *J. Dent.* **14**, 231 (1986).
46. A. D. Wilson and J. W. McLean, *Glass-Ionomer Cement*, Quintessence Publishing Co., Chicago, 1988.
47. J. W. McLean, *Br. Dent. J.* **164**, 293 (1988).
48. G. J. Mount, in *The Clinician's Handbook*, Martin Dunitz, London, 1989.
49. A. Maldonado, M. L. Swarz, and R. W. Phillips, *J. Am. Dent. Assoc.* **96**, 785 (1978).

50. S. Crisp, B. G. Lewis, and A. D. Wilson, *J. Dent.* **7**, 304 (1979).

51. R. G. Mill and A. D. Wilson, *J. Dent. Res.* **67**, 1446 (1988).

52. J. W. Nicholson and co-workers, *J. Dent. Res.* **67**, 1451 (1988).

53. J. W. McLean, A. D. Wilson, and H. J. Prosser, *J. Prosthet. Dent.* **52**, 175 (1984).

54. P. Hotz and co-workers, *Br. Dent. J.* **141**, 41 (1977).

55. D. P. Powis and co-workers, *J. Dent. Res.* **61**, 1416 (1982).

56. B. L. Dahl and G. Oilo, *Dent. Mater.* **2**, 17 (1986).

57. Y. E. Y. Aboush and C. B. G. Jenkins, *J. Dent.* **15**, 147 (1986).

58. J. W. McLean and co-workers, *Br. Dent. J.* **158**, 410 (1985).

59. F. I. Magnum, E. A. Berry, and U. K. Parikh, *J. Am. Dent. Assoc.* **120**, 535 (1990).

60. J. E. Rusz and co-workers, *Dent. Mater.* **8**, 31 (1992).

61. J. E. McKinney, J. M. Antonucci, and N. W. Rupp, *J. Dent. Res.* **66**, 1134 (1987).

62. H. J. Prosser, D. R. Powis, and A. D. Wilson, *J. Dent. Res.* **65**, 146 (1986).

63. J. W. McLean and O. Gasser, *Quintessence Int.* **16**, 333 (1985).

64. J. E. McKinney and J. M. Antonucci, *J. Dent. Res.* **66**, (1980).

65. J. M. Antonucci, J. E. McKinney, and J. W. Stansbury, *Trans. 13th Ann. Mtg. Soc. Biomater.*, 1987, p. 225.

66. J. M. Antonucci, *Trends and Techniques* **5**, 4 (1988).

67. J. M. Antonucci and J. W. Stansbury, *J. Dent. Res.* **68**, 251 (1989).

68. R. S. Mathis and J. L. Ferracane, *Dent. Mater.* **5**, 355 (1989).

69. J. W. McLean, *J. Am. Dent. Assoc.* **120** 43 (1990).

70. J. J. Simmons, *J. Am. Dent. Assoc.* **120** 49 (1990).

71. **U.S. Pat. 4,527,979** (July 9, 1985), J. W. McLeonard and O. Gasser (to ESPE, Fabrik Pharmazeutischer Praeparate, GmbH).

72. J. E. McKinney, J. M. Antonucci, and N. W. Rupp, *J. Dent. Res.* **67**, 83 (1988).

73. S. B. Mitra, *J. Dent. Res.* **70**, 72 and 75 (1991).

74. S. B. Mitra, *J. Dent. Res.* **70**, 72 (1991).

75. *Ibid.*, 75 (1991).

76. A. Sugawara and co-workers, *J. Nihon Sch. Dent.* **31**, 372 (1989).

77. G. V. Balck, *Operative Dentistry*, Vol. 3, 7th ed., Medico Dental Publishing Co., Chicago, 1936, p. 233.

78. D. A. Wallace and H. C. Hanson, *J. Am. Dent. Assoc.* **26**, 1536 (1939).

79. W. Harvey and W. J. Pitch, *Br. Dent. J.* **80**, 1 (1946).

80. E. J. Molnar and E. W. Skinner, *J. Am. Dent. Assoc.* **29**, 744 (1942).

81. H. I. Copeland and co-workers, *J. Dent. Res.* **34**, 740 (1955).

82. J. M. Antonucci and co-workers, *2nd World Congress on Biomaterials, Trans. 10th Ann. Mtg. Soc. Biomater.*, 1984, p. 137.

83. M. D. Jendresen, M. L. Swartz, and R. D. Norman, *J. Prostb. Dent.* **21**, 176 (1969).

84. M. D. Jendresen and R. W. Phillips, *J. Prostb. Dent.* **21**, 300 (1969).

85. G. M. Brauer, R. McLaughlin, and E. Huget, *J. Dent. Res.* **47**, 622 (1968).

86. A. D. Wilson, in J. A. von Fraunhoffer, ed., *Scientific Aspects of Dental Materials*, Butterworths, London 1975, Chapt. 4–5.

87. D. F. Vieira, *J. Prosthet. Dent.* **9**, 70 (1959).

88. M. T. Tyas and H. J. Wilson, *Br. Dent. J.* **129**, 461 (1970).

89. S. Olson, B. Bergman, and M. Bergman, *Swed. Dent. J.* **6**, 177 (1982).

90. G. M. Brauer, E. E. White, and M. G. Mashonas, *J. Dent. Res.* **37**, 547 (1958).

91. G. M. Brauer, L. Simon, and L. Sangermono, *J. Dent. Res.* **41**, 1096 (1962).

92. S. Civjan and G. M. Brauer, *J. Dent. Res.* **43**, 281 (1964).

93. G. M. Brauer and J. W. Stansbury, *J. Dent. Res.* **63**, 137 (1984); *Ibid.*, p. 1315.

94. J. W. Stansbury and G. M. Brauer, *J. Biomed. Mater. Res.* **19**, 715 (1985).

95. G. M. Brauer, J. W. Stansbury, and D. Flowers, *Dent. Mater.* **2**, 21 (1986).

96. **U.S. Pat. 3,047,408** (July 31, 1962), E. W. Dougherty (to L. D. Caulk Co.).

97. F. J. Fisher, *Br. Dent. J.* **143**, 231 F(1977).

98. W. L. Lang, *J. Dent. Res.* **65**, 175 (1986).

99. N. S. Seale and H. R. Stanley, *J. Dent. Res.* **65**, 772 (1986).

100. P. J. Schouboe, G. C. Paffenbarger, and W. T. Sweeney, *J. Am. Dent. Assoc.* **52**, 584 (1956).

101. C. L. Hadden, *Chem. Ind.* **21**, 190 (1944).

102. *Dental Phosphate-Bonded Casting Investments, International Standard ISO 9694*, American National Standards Institute, New York.

103. D. B. Mahler and K. Asgarzadeh, *J. Dent. Res.* **32**, 354 (1953).

104. K. D. Jorgensen, *Odontol. Tidskr.* **61**, 305 (1953).

105. C. W. Fairhurst, *J. Dent. Res.* **39**, 812 (1960).

106. *Guide to Dental Materials and Devices*, 8th ed., American Dental Assoc., Chicago, 1976–1978, p. 110.

107. D. B. Mahler and A. B. Ady, *J. Dent. Res.* **39**, 578.

108. R. H. Volland and G. C. Paffenbarger, *J. Am. Dent. Assoc.* **19**, 185 (1932).

109. R. Earnshaw, *Aust. Dent. J.* **14**, 264 (1969).

110. *Guide to Dental Materials and Devices*, 8th ed., American Dental Assoc., Chicago, 1976–1978, p. 112.

111. R. Earnshaw, in W. J. O'Brian ed., *Dental Materials, Properties and Selection*, Quintessence Publishing Co., Chicago, 1989.

112. W. J. O'Brien and J. P. Nielsen, *J. Dent. Res.* **38**, 541 (1959).

113. **U.S. Pat. 1,909,008** (May 16, 1933), C. H. Prange (to Austenal Laboratories, Inc.).

114. C. P. Mabie, *J. Dent. Res.* **52**, 758 (1973).

115. **U.S. Pat. 2,222,781** (Nov. 26, 1940), T. E. Moore (to Ransom and Randolph Co.).

116. C. P. Mabie, *J. Dent. Res.* **52**, 96 (1973).

117. R. J. Schnell, G. Mumford, and R. W. Phillips, *J. Prosthet. Dent.* **13**, 324 (1963).

118. **U.S. Pat. 1,708,436** (Apr. 9, 1929), L. J. Weinstein.

119. **U.S. Pat. 1,924,874** (Aug. 29, 1933), T. E. Moore (to Ransom and Randolph Co.).

120. N. O. Taylor, G. C. Paffenbarger, and W. T. Sweeney, *J. Am. Dent. Assoc.* **17**, 2266 (1930).

121. J. S. Shell, *Dent. Surv.* **36**, 616 and 770.

122. *Dental Gypsum-Bonded Casting Investments for Gold Alloys, International Standard ISO 7490*, American National Standards Institute, New York.

123. R. Earnshaw, *J. Dent. Res.* **67**, 1366 (1988).

124. D. R. Davis, *J. Prosthet. Dent.* **57**, 362 (1987).

125. L. Forsten, *Scand. J. Dent. Res.* **98**(2), 179 (1990).

126. M. Ohashi and G. C. Paffenbarger, *J. Nihon Univ. Sch. Dent.* **11**, 109 (1969).

127. L. C. Dirksen, *J. Am. Dent. Assoc.* **26** 273 (1939).

128. J. W. McCrorie, *Br. Dent. J.* **132** 189 (1972).

129. J. F. Bates and A. G. Knapton, *Int. Metall. Rev.* **22**, 39 (Mar. 1977).

130. P. Reithe, *Tsch. Zahnärztl. Z.* **21**, 301 (1966).

131. D. B. Mahler and J. van Eysden, *J. Dent. Res.* **48**, 501 (1969).

132. D. B. Mahler and co-workers, *J. Dent. Res.* **49**, 1452 (1970).

133. D. B. K. Innes and W. V. Youdelis, *J. Can. Dent. Assoc.* **29**, 587 (1963).

134. D. B. Mahler, L. G. Terkla, and J. van Eysden, *J. Dent. Res.* **52**, 823 (1973).

135. K. Asgar, *J. Dent. Res.* **53**, special issue, abst. 23 (1974).

136. J. W. Osborne and R. D. Norman, *Dent. Mat.* **6**, 189 (1990).

137. L. B. Johnson, *J. Biomed. Mater. Res.* **4**, 269 (1970).

138. R. Nadal, R. W. Phillips, and M. L. Swartz, *J. Am. Dent. Assoc.* **63**, 488 (1961).

139. R. M. Waterstrat, *J. Am. Dent. Assoc.* **78**, 536 (1969).

140. J. P. Nielsen and J. T. Tuccillo, *J. Dent. Res.* **45**, 964 (1966).

141. D. L. Smith and co-workers, *J. Dent. Res.* **49**, 283 (1970).

142. S. Civjan and G. M. Brauer, *J. Dent. Res.* **43**, 281 (1964).

143. D. C. Smith, *Br. Dent. J.* **124**, 381 (1968).

144. *Metals Handbook*, 8th ed., American Society for Metals, Metals Park, Ohio, 1961, p. 1189.

145. J. N. Nally, *Int. Dent. J.* **18**, 309 (1968).

146. S. Civjan, D. F. Huget, and J. E. Marsden, *J. Am. Dent. Assoc.* **85**, 1309 (1972).

147. S. Civjan and co-workers, *J. Am. Dent. Assoc.* **90**, 659 (1975).

148. E. F. Huget and S. Civjan, *J. Am. Dent. Assoc.* **89**, 383 (1974).

149. R. W. Phillips, *Skinner's Science of Dental Materials*, 7th ed., W. B. Saunders Co., Philadelphia, 1973, p. 565.

150. R. L. Coleman, *Burr. Stand. U.S. J. Res.* **1**, 867 (1928).

151. **U.S. Pat. 1,909,008** (May 16, 1933), C. H. Prange (to Austenal Laboratories, Inc.).

152. **U.S. Pat. 1,956,278** (Apr. 24, 1934), R. W. Erdle and C. H. Prange (to Austenal Laboratories, Inc.).

153. **U.S. Pat. 1,958,446** (May 15, 1934), C. H. Prange (to Austenal Laboratories, Inc.).

154. H. F. Morris and K. Asgar, *J. Pros. Dent.* **33**, 36 (1975).

155. J. Jahtzen, *Quintessence Int.* **1**, 1 (1970).

156. J. F. Bates and A. G. Knapton, *Int. Metall. Rev.* **22**, 39 (1977).

157. R. W. Phillips, *Skinner's Science of Dental Materials*, 7th ed., W. B. Saunders Co., Philadelphia, 1973, p. 593.

158. R. Earnshaw, *Aust. Dent. J.* **3**, 159 (1958).

159. G. T. Eden and co-workers, *J. Dent. Res.* **56**, special issue, abst. 648 (1977).

160. J. P. Moffa, in T. M. Valega, Sr. ed., *Alternatives to Gold Alloys*, conference proceedings, Jan. 24–26, 1977, DHEW publication, Bethesda, Md., (NIH) no. 77-1227, 1977.

161. E. F. Huget, J. M. Vilca, and R. M. Wall, *J. Dent. Res.* **56**, special issue B, abst. 642 (1977).

162. J. F. Bates and A. G. Knapton in Ref. 77.

163. J. P. Moffa and W. A. Jenkins in Ref. 54, p. 652; N. K. Sarkar and E. H. Greener, *Materials Engineering Conference*, Oct. 1973, Chicago, American Society for Metals, Metals Park, Ohio.

164. O. E. Beder and W. J. Ploger, *Oral Surg. Oral Med. and Oral Path.* **12**, 787 (1959).

165. T. M. Valega, Sr. ed., *Alternatives to Gold Alloys*, conference proceedings, Jan. 24–26, 1977, DHEW publication, Bethesda, Md., (NIH) no. 77-1227, 1977, pp. 224–233.

166. K. Ida and co-workers, *Dent. Mat. J.* **1**, 8 (1982).

167. H. J. Mueller, A. A. Giuseppetti, and R. M. Waterstrat, *J. Dent. Res.* **69** special issue, 2072 (1990).

168. Y. Tamaki and co-workers, *J. Dent. Res.* **70**, special issue, abst. 1755 (1991).

169. R. J. Baez, T. Nonaka, and R. Blackman, *J. Dent. Res.* **70**, special issue, abst. 1757 (1991).

170. A. S. Algabab and co-workers, *Northwestern Dent. Res.* **2**, 21 (1991).

171. A. R. Hruska, *Quintessence Int.* **18**, 683 (1987).

172. A. R. Hruska, *Quintessence Int.* **15**, 198 (1989).

173. F. Watari, *Dent. Mat. J.* **8**, 83 (1989).

174. A. Kuroiwa and co-workers, *Dent. Mat. J.* **9**, 279 (1990).

175. S. P. Thompson and co-workers, *J. Dent. Res.* **71**, special issue, abst. 1418 (1992).

176. N. Morita, *Dent. Mat. J.* **9**, 218 (1990).

177. Y. Tamaki and co-workers, *Dent. Mat. J.* **8**, 103 (1980).

178. A. H. L. Tjan and co-workers, *J. Prosthet. Dent.* **66**, 157 (1991).

179. Ref. 129, p. 39.

180. *Brazing Manual*, American Welding Society, United Engineering Center, New York, 1955.

181. J. F. Johnston, G. Mumford, and R. W. Dykema, *Modern Practice in Dental Ceramics*, W. B. Saunders Co., Philadelphia, 1967.

182. *Specification B 32-60aT*, ASTM Standards, 1992.

183. *Soldering Manual*, American Welding Society, United Engineering Center, New York, 1959.

184. *Metals Handbook*, 8th ed., American Society of Metals, Metals Park, Ohio, 1961, pp. 1188–1189.

185. R. W. Phillips, *Skinner's Science of Dental Materials*, 7th ed., W. B. Saunders Co., Philadelphia, 1973, Chapt. 31–33.

186. E. M. Wise, *Gold: Recovery, Properties and Applications*, D. Van Nostrand Co., Inc., New York, 1964.

187. T. M. Valega, Sr., ed., in Ref. 165, pp. 224–233.

188. E. H. Stade, M. H. Reisbick, and J. D. Preston, *J. Prosthet. Dent.* **34**, 527 (1975).

189. R. G. Craig, in B. R. Land and C. C. Kelsey, eds., *International Prosthodontics Workshop on Complete Denture Occlusion*, University of Michigan Press, Ann Arbor, Mich., 1973, pp. 205–236.

190. K. K. Kapur and S. D. Somani, *J. Prosthet. Dent.* **14**, 687 (1964).

191. R. V. Reitz, J. L. Sanders, and B. Levin, *Quintessence Int.* **16**, 547 (1985).

192. J. L. Sanders, B. Levin, and P. V. Reitz, *Quintessence Int.* **18**, 453 (1987).

193. D. A. L. Doorri and co-workers, *Dent. Mater.* **4**, 25 (1988).

194. V. T. Truong and F. G. V. Thomas, *Aust. Dent. J.* **33**, 201 (1988).

195. S. Winkler, J. P. Powell, and H. R. Ortman, *Dent. Dig.* **77**, 460 (1971).

196. J. B. Woelfel, G. C. Paffenbarger, and W. T. Sweeney, *J. Am. Dent. Assoc.* **70**, 1170 (1965).

197. W. E. Mowery and co-workers, *J. Am. Dent. Assoc.* **57**, 345 (1958).

198. W. L. Shepard, *J. Prosthet. Dent.* **19**, 561 (1968).

199. J. B. Woelfel, *Dent. Clin. North Am.* **15**, 67 (1971).

200. N. W. Rupp, G. C. Paffenbarger, and R. L. Bowen, *J. Am. Dent. Assoc.* **83**, 601 (1971).

201. J. B. Woelfel, G. C. Paffenbarger, and W. T. Sweeney, *J. Am. Dent. Assoc.* **61**, 413 (1960).

202. *Dentsply TRLAD Denture System, Pamphlet 3542C*, Dentsply Int., York, Pa., 1983.

203. H. Wolfe, *Dent. Tech.* **37**, 6 (1984).

204. J. B. Woelfel, G. C. Paffenbarger, and W. T. Sweeney, *J. Am. Dent. Assoc.* **67**, 489 (1963).

205. H. H. Chandler, R. L. Bowen, and G. C. Paffenbarger, *J. Biomed. Mater. Res.* **5**, 245 (1971).

206. Council on Dental Materials, Instruments, and Equipment, *J. Am. Dent. Assoc.* **102**, 347 (1981).

207. Ref. 205, p. 335.

208. *Dentist's Desk Reference: Materials, Instruments and Equipment*, 1st ed., American Dental Assoc., Chicago, 1981, p. 97.

209. R. Chapman and co-workers, *LADR Abst.* 959 (1982).

210. M. Braden and co-workers, *Br. Dent. J.* **164**, 109 (1988).

211. M. Braden, *Adv. in Dent. Res.* **2**, 93 (1988).

212. C. S. Bayne, *Adv. in Dent. Res.* **98**, (1988).

213. J. W. Stamford, C. L. Burns, and G. C. Paffenbarger, *J. Am. Dent. Assoc.* **61**, 307 (1955).

214. P. S. Wright, *J. Dent.* **4**, 247 (1976).

215. W. J. O'Brien, J. Herman, and J. H. Shepherd, *J. Biomed. Mat. Res.* **6**, 15 (1972).

216. J. F. McCabe, *J. Oral Rehab.* **3**, 273 (1976).

217. E. Makila and O. Honka, *J. Oral Rehab.* **6**, 199 (1979).

218. L. Gittleman and co-workers, in *Biomedical Engineering IV: Recent Developments*, Pergamon Press, New York, 1985, pp. 141–144.

219. M. Braden, *J. Dent. Res.* **49**, 145 (1970).

220. M. Braden, in *Scientific Aspects of Dental Materials*, Butterworths, London, 1975, pp. 448–458.

221. *Ibid.*, pp. 439–448.

222. R. J. Michl, *Quintessence Int.* **3**, 29 (1978).

223. J. C. Hickey and co-workers, *J. Am. Dent. Assoc.* **74**, 735 (1967).

224. J. M. Stenger and co-workers, *J. Am. Dent. Assoc.* **69**, 273 (1964).

225. R. E. Going, R. E. Loehman, and M. S. Chan, *J. Am. Dent. Assoc.* **89**, 132 (1973).

226. S. J. Chaconas, A. A. Caputo, and N. K. Bakke, *J. Dent. Res.* **63**, 1488 (1984).

227. D. H. Lewis, T. E. Fischer, and A. Koran, in *Restorative Dental Materials*, Quintessence Publishing Co., Chicago, 1985, pp. 214–241.

228. J. M. Antonucci, *Transactions of the Academy of Dental Materials; Proceedings of the Conference on Materials Research in Maxillofacial Prosthetics*, 1992, pp. 136–155.

229. G. M. Brauer and J. M. Antonucci, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, Vol. 4, 2nd ed., Wiley-Interscience, New York, 1986, p. 705.

230. **Brit. Pat. 252,112** (May 14, 1925), A. Poller (to DeTrey Bros. Ltd.).

231. **U.S. Pat. 1,672,776** (June 5, 1928), A. Poller (to DeTrey Bros. Ltd.).

232. **U.S. Pat. 2,249,694** (July 5, 1939), S. W. Wilding (to the Amalgamated Dental Co.).

233. **U.S. Pat. 2,434,005** (Jan. 6, 1948), S. E. Noyes.

234. **U.S. Pat. 2,454,709** (Nov. 23, 1948), E. J. Molnar (to Montclair Research Corp.).

235. W. D. Cook, *J. Biomed. Mater. Res.* **16**, 315 (1982).

236. *Ibid.*, 331.

237. *Ibid.*, 345.

238. *Ibid.*, 711–713.

239. R. G. Craig, in *Advances in Dental Research*, 1988, pp. 51–64.

240. **Fr. Pat. 1,423,660** (Jan. 7, 1966), (to ESPE, Fabrik Pharmazeutischer Praeparate, GmbH).

241. **U.S. Pat. 3,453,242** (July 1, 1969), W. Schmitt and R. Purrmann (to ESPE, Fabrik Pharmazeutischer Praeparate, GmbH).

242. M. Braden, B. E. Causton, and R. L. Clarke, *J. Dent. Res.* **51**, 889 (1972).

243. R. G. Craig, *J. Am. Dent. Assoc.* **95**, 127 (1977).

244. R. G. Craig and P. H. Hare, *J. Prosthet. Dent.* **63**, 16 (1990).

245. **U.S. Pat. 3,066,112** (Nov. 27, 1962), R. L. Bowen (to U.S. Dept. of Commerce).

246. **U.S. Pat. 3,179,623** (Apr. 20, 1965), R. L. Bowen (to U.S. Dept. of Commerce).

247. M. L. Swartz, R. W. Phillips, and B. F. Rhodes, *J. Dent. Res.* **61**, 27 (1982).

248. N. O. Salaki and co-workers, *Br. Dent. J.* **146**, 375 (1979).

249. M. Berge, *Dent. Mater.* **5**(2) 77 (1989).

250. S. L. Wendt, Jr., and co-workers, *J. Am. Dent. Assoc.* **120**, 177 (1990).

251. G. M. Brauer, in R. B. Graig, ed., *Polymers in Dentistry, Dental Material Reviews*, University of Michigan School of Dentistry, Ann Arbor, Mich., 1977, p. 49.

252. R. L. Bowen, *J. Am. Dent. Assoc.* **66**, 57 (1963).

253. R. W. Phillips, *Science of Dental Materials*, W. B. Saunders Co., Philadelphia, 1991, pp. 217–235.

254. I. E. Cueto and M. G. Buonocore, *J. Am. Dent. Assoc.* **75**, 121 (1967).

255. M. G. Buonocore, *J. Am. Dent. Assoc.* **82**, 1090 (1971).

256. W. P. Rock, *Br. Dent. J.* **134**, 193 (1973).

257. R. J. McCune and J. F. Cvar, *Int. Assoc. Dent. Res. Program and Abstracts of Papers, abst. 745*, Chicago, International Association for Dental Research, Washington, D.C., 1971.

258. R. J. Simonsen, *Quintessence Int.* **20**(2) 75 (1989).

259. E. P. Pleuddemann, *Silane Coupling Agents*, Plenum Press, New York, 1982, pp. 111–139.

260. E. P. Pleuddemann, *J. Adhesion* **2**, 184 (1970).

261. I. Stengel, D. Nathanson, and C. S. Hsu, *J. Dent. Res.* **66**, 1460 (1987).

262. G. J. Livaditis and V. P. Thompson, *J. Prosthet. Dent.* **47**, 52 (1982).

263. G. J. Livaditis, *J. Prosthet. Dent.* **56**, 181 (1986).

264. W. A. Wiltshire, *Quintessence Dent. Tech.* **10**, 277 (1986).

265. Y. Kondo and co-workers, *Shika Zairyo Kikai* (1), (1989).

266. I. Harashima and co-workers, *Dent Mater. J.* **7**(2) 141 (1988).

267. L. T. Tenjoma, *Int. J. Prosthodont.* **3**(1) 78 (1990).

268. M. G. Buonocore, *J. Dent. Res.* **34**, 849 (1955).

269. D. H. Anthony, *M.S. thesis*, University of Michigan School of Dentistry, 1961.

270. A. Watanabe and N. Nakabayashi, *Tokyo Ika Shika Daigaku Iyo Kizai Kenkyusho Hokou*, **21**, 17 (1987).

271. R. L. Bowen, *J. Dent. Res.* **44**, 895, 903, 906 (1969).

272. D. N. Misra and A. D. Johnston, *J. Biomed. Mater. Res.* **21**, 1329 (1987).

273. D. C. Smith, *Br. Dent. J.* **125**, 381 (1968).

274. A. Jurecik and C. W. Fairhurst, *Int. Assoc. Dent. Res., Program and Abstracts of Papers, abst. 379*, Las Vegas, Nev., International Association for Dental

Research, Washington, D.C., 1972.

275. D. R. Beech, *Arch. Oral Biol.* **17**, 907 (1972).

276. S. Crisp and A. D. Wilson, *J. Dent. Res.* **53**, 1408 (1974).

277. S. Crisp, I. W. Pringuer, and A. D. Wilson, *J. Dent. Res.* **53**, 1414 (1974).

278. S. Crisp and A. D. Wilson, *J. Dent. Res.* **53**, 1420 (1974).

279. J. W. McLean and co-workers, *Br. Dent. J.* **158**, 410 (1985).

280. Ref. 178, p. 157.

281. J. Arends and co-workers, *Quintessence Int.* **21**, 541 (1990).

282. M. E. Jensen and co-workers, *Am. J. Dent.* **3**(2), 75 (1990).

283. M. L. Capilouto and co-workers, *Caries Res.* **24** 6, 441 (1990).

284. Ref. 125, p. 179.

285. L. Forsten, *Scand. J. Dent. Res.* **99**(3), 241 (1991).

286. A. Tewari and co-workers, *J. Indian Soc. Pedod. Prev. Dent.* **8**(1), 28 (1991).

287. H. S. Horowitz and co-workers, *J. Public Health Dent.* **51**(1), 20 (1991).

288. *Physicians Desk Reference*, 45th ed., Medical Economics Co. Inc., Oradell, N.J., 1991.

289. Center for Disease Control, *JAMA*, **266**, 1061 and 1066 (1991).

290. M. Jarco, *J. Mater. Sci.* **11**, 2027 (1976).

291. R. A. White, J. M. Webber, and E. W. White, *Science* **176**, 187 (1972).

292. R. E. Holmes and H. K. Hagler, *Plast. Reconstr. Surg.* **81**, 662 (1988).

293. K. de Groot and co-workers, in T. Yamamuro, L. L. Hench, and J. Wilson, eds., *Handbook of Bioactive Ceramics*, Vol. 2, CRC Press Inc., Boca Raton, Fla., 1990, p. 3.

294. R. Z. LeGeros, *Adv. Dent. Res.* **2**, 164 (1988).

295. T. Yamamuro, L. L. Hench, and J. Wilson in Ref. 293.

296. R. Z. LeGeros and G. Daculsi, in Ref. 293.

297. L. C. Chow, *J. Ceramic Soc. Japan* **99**, 954 (1991).

298. A. A. Mirtchi, L. Lemaitre, and N. Terao, *Biomater.* **10**, 475 (1989).

299. H. Monma, M. Goto, and T. Kohmura, *Gypsum & Lime*, **188**, 11 (1984).

300. W. E. Brown and L. C. Chow, in P. W. Brown, ed., *Cements Research Progress*, American Ceramics Society, Westerville, Ohio, 1986, p. 352.

301. L. C. Chow and co-workers, *Matl. Res. Symp. Proc.* **179**, 3 (1991).

302. Y. C. Hong and co-workers, *J. Biomed. Matls. Res.* **24**, 485 (1991).

303. P. D. Costantino and co-workers, *Arch. Otolaryngol. Head Neck Surg.* **117**, 379 (1991).

304. C. D. Friedman and co-workers, *Arch. Otolaryngol. Head Neck Surg.* **117**, 385 (1991).

305. A. Sugawara and co-workers, *J. Endodon.* **16**, 162–165 (1990).

306. Y. C. Hong and co-workers, *J. Biomed. Matls. Res.* **25**, 485–498 (1991).

307. A. Sugawara and co-workers, *Tokyo Shika Ishikai Zasshi*, **38**, 348–354 (1990).

308. C. D. Friedman and co-workers, *Arch. Otolaryngol. Head Neck Surg.* **117**, 385–389 (1991).

309. P. D. Constantino, C. D. Friedman, R. Jones, L. C. Chow, and G. A. Sisson, *Plast. Reconstr. Surg.* **90**, 174–191 (1992).

310. National Institutes of Health, *Consensus Development Conference Statement on Dental Implants*, **7** (1988).

311. T. Albrektsson and co-workers, *Int. J. Maxillofac. Impl.* **1**, 11, (1986).

312. L. B. Shulman, *NIH Consensus Development Conference Statement on Dental Implants* **7**, 43, (1988).

313. H. Hahn, personal communication, Artech Reliance Corp., Chantilly, Va., 1992.

314. P. I. Branemark, *J. Prosthet. Dent.* **50**, 399–410 (1983).

315. *Code of Fed. Reg., pt. 872.3640*, Vol. 21, Off. Fed. Regist., Nat. Archives and Rec. Admin., Washington, D.C., Apr., 1990.

316. A. C. Fraker and co-workers, in M. J. Fagan, Jr. and co-workers, eds., *Implant Prosthodontics: Surgical and Prosthetic Techniques for Dental Implants*, Yearbook Medical Publishers, Littleton, Mass., 1991, pp. 293–304.

317. D. Adams and D. F. Williams, *A Review of Dental Implants*, Dental Update, Upstate Publications, Ltd., London, 1985, p. 480.

General Reference

G. C. Paffenbarger, J. A. Tesk, and W. E. Brown, *J. Am. Dent. Assoc.* **116**, 83 (1985).

John A. Tesk
Joseph M. Antonucci
Fred C. Eichmiller
J. Robert Kelly
Nelson W. Rupp
Richard W. Waterstrat
Anna C. Fraker
Lawrence C. Chow
Laurie A. George
Jeffrey W. Stansbury
Edward E. Parry
National Institute of Standards and Technology (NIST)

DENTIFRICES

Dentifrices are compositions designed to be used with a toothbrush. The two essential functions of dentifrices are to remove dental stains and to introduce a fresh, cool, pleasant, and clean feeling into the oral cavity. In the 1960s, a third function was introduced, delivery to the dentition of the therapeutic chemical agent fluoride. Subsequently, agents offering specific therapeutic and nontherapeutic benefits were added to dentifrices. The three forms of marketed dentifrices are pastes (toothpastes), liquid concentrates, and powders. Pastes account for an overwhelming portion of the dentifrice market. Liquid concentrates and powders account for only a minor portion because they have disadvantageous dispensing characteristics and they do not provide a means to accurately and consistently dispense a dose of active agent.

Dentifrices and dental rinses have many common purposes and ingredients, thus dental rinses are discussed briefly herein.

Dental Plaque. Dental plaque, a mass of bacteria that develops around teeth in all people, is the primary cause of dental caries and diseases of the periodontium. Following dental prophylaxis, a tooth becomes covered with a thin film, or pellicle, selectively deposited from the saliva. Bacteria in the oral cavity attach to the pellicle. The early colonizers are predominantly gram-positive bacteria. The bacterial population increases and undergoes a predictable pattern of change as it approaches maturity. Supragingival and subgingival plaque masses are distinguishable. Supragingival plaque matures in about 10 to 12 days, at which time calcification of the mass may be observed (calculus, tartar). The maturation process may be altered by intervening events, such as toothbrushing. The microbial population may generate acids, which cause dental caries, and toxins, which cause irritation of the gingiva, depending on environmental and other factors.

Mechanical removal of plaque is the most effective measure against plaque-caused diseases, dental caries, and periodontal diseases. Even before the advent of fluoride treatments, it was assumed that a clean tooth does not decay. A toothbrush is effective in removing dental plaque and, for those individuals who optimize its use, it usually can adequately control plaque. Despite the proven efficacy of mechanical plaque removal, the amount of patient involvement is such that only about 30% of the population in developed countries and considerably less in undeveloped countries can be expected to adequately remove plaque (1). Hence, supplementary measures such as dentifrices and dental rinses are necessary.

Studies have shown that compliance with toothbrushing would be severely impaired if the brushing regimen did not include dentifrices. Thus, the role of dentifrices in plaque control is to enhance the toothbrushing experience and thereby enhance the results. Dentifrices have little direct effect on plaque removal or accumulation unless they contain agents that have specific antiplaque properties.

Toothpastes

Toothpastes are packaged in flexible tubes, other flexible containers, and mechanically operated pump dispensers. They are usually extruded as cylindrical ribbons of a cohesive, smooth paste, approximately 2.54 cm in length and weighing approximately 1.5 g. New or modified dispensing devices are continually introduced to increase consumer interest.

The addition of therapeutic or cosmetic agents to dentifrices has paralleled advances in knowledge about factors affecting the human dentition. Agents added to dentifrices can act directly on the host tooth structure or on specific oral accumulations, for example, the principal action of fluoride is on the tooth enamel. The primary action of an abrasive, however, is on an accumulated stained pellicle. Oral accumulations of interest to preventive dentistry are dental pellicles, dental plaque, dental calculus (tartar), microbial populations responsible for oral malodor, and oral debris (food residues, leukocytes, etc). Plaque is most important because of its potential to do harm.

General Toothpaste Formulation.

The key functions of dentifrices, to remove dental stains and to freshen the oral cavity, are accomplished by abrasive cleaning and the masking or elimination of unpleasant oral odors. Materials designed to deliver antitartar, antiplaque, or anticaries benefits must be compatible with the ability of the dentifice to fulfill those two functions.

Toothpaste contains an abrasive (qv), flavor, a humectant system, a surfactant, a binding and thickening agent, color, and one or more therapeutic or cosmetic agents.

Abrasive. Dentifrices have the unique ability to remove extrinsic tooth stains, which are caused by agents such as berries, tea, smoking, antibiotics, and certain bacteria as they attach to the dental pellicle. These stains can be removed only by abrasive cleaning; a toothbrush alone is not adequately effective. It has been shown that only 4% of a test population were able to maintain their teeth in an acceptably stain-free state without an abrasive and that 18% of the population were "heavy" stainers (2). However, colored materials found in dental plaque are removable without abrasives.

Many abrasive materials have been incorporated into dentifrices to remove stained pellicles. These are generally particles of insoluble inorganic materials and are characterized for dentifrice application by five interdependent parameters: particle size, shape, hardness, toughness, and chemical inertness. Particle size is limited by safety and comfort. The larger the particle, the more it abrades and the more readily it can be detected in the mouth as a gritty foreign material. Most abrasives for dentifrices have an average particle diameter of about 3 to 12 μm. Particles of 30 μm or larger are detected as gritty. The particle shape should be almost spherical, with no protuberances. Sharp points scratch the surface of the tooth. Hardness is governed by the hardness of dentin; on Mohs' scale the hardness of dentin is 2–2.5 and the hardness of dental enamel is 4–5. (Hardness values of talc [14807-96-6] and quartz [14808-60-7] are 1 and 7, respectively, on Mohs' scale). A hardness of about 2.5 for the abrasive material is adequate to assure that the stained pellicle is safely removed. Toughness is an expression of the tendency of the particle to break down under the force of toothbrushing. Some materials appear to be too hard for use as an abrasive. These can be used, however, if they break down during toothbrushing to particles too small to harm dentin or enamel.

An abrasive is usually chemically inert, neither interacting with other dentifrice ingredients nor dissolving in the paste or the mouth. Substances used as dentifrice abrasives include amorphous hydrated silica, dicalcium phosphate dihydrate [7789-77-7], anhydrous dicalcium phosphate [7757-93-9], insoluble sodium metaphosphate [10361-03-2], calcium pyrophosphate [35405-51-7], α-alumina trihydrate, and calcium carbonate [471-34-1]. These materials are usually synthesized to specifications for purity, particle size, and other characteristics; naturally occurring minerals are used infrequently. Sodium bicarbonate [144-55-8] and sodium chloride [7647-14-5] have also been employed as dentifrice abrasives.

Methods have been developed to quantify the abrasiveness of dental abrasives. Numerical values are assigned based on the amount of dentin removed by toothbrushing using a standardized slurry of abrasive or total dentifrice under standardized conditions, and comparison with a reference abrasive. One popular measure is the radioactive dentin abrasion value (RDA) (3). This value is obtained by subjecting an excised bovine tooth to a neutron flux, mounting segments of the irradiated tooth dentin in a special brushing machine, and brushing under standardized conditions. The amount of ³²P removed is compared with the amount of ³²P removed using a reference calcium pyrophosphate abrasive set at 500. The RDA of the total finished product is also measured, because the apparent abrasiveness of an abrasive can be modified by the presence of other ingredients of the formulation. Safe and effective RDA values are in the range of about 50 to 250.

Flavor. Dentifrices are used to refresh the oral cavity. Flavor oils and other flavoring materials are key to that function (see FLAVORS AND SPICES). Generally recognized as safe (GRAS) flavors or flavors from approved lists are used. The most popular flavors are peppermint [8006-90-4], spearmint [8008-79-5], cinnamon [8006-79-9], and mixtures of these. Menthol is a principal constituent of the mint flavors and a source of refreshment and coolness.

The concentration of the flavor in the dentifrice and the nature of substances added to bring out specific flavor notes and thereby make the flavor unique are significant concerns. The flavor must not be excessive; it must not burn too strongly. Also, the flavor must not be a sensitizer. Synthetic sweeteners are usually added, although regulatory concerns limit their selection; for example, cyclamate [100-88-9] is not used in the United States.

Humectant System. The humectant system comprises one or more liquids in which the other toothpaste ingredients are dispersed to provide a stable paste. The liquids are usually glycerol and 70% aqueous sorbitol. Propylene glycol [57-55-6] and polyethylene glycols are used occasionally. The liquids are used alone or in combination. Their water activity is such that they do not support bacteriological growth. The humectant system does not crystallize, and thus the final formulation does not dry out or cause the tube cap to lock to the nozzle.

Surfactant. The primary purpose of a surfactant in toothpaste is to create a foam while the teeth are brushed. This foam provides an enjoyable sensation. Secondly, the surfactant helps remove material dislodged by the toothbrush, and it may have minor effects on plaque accumulation (see SURFACTANTS).

The surfactant most commonly used is the anionic detergent sodium lauryl sulfate. Other surfactants that have been used include sodium dodecylbenzene sulfonate [25155-30-0], sodium N-lauroyl sarcosinate or Gardol [137-16-6], and sodium cocomonoglyceride sulfonate [3694-90-4]. Cationic and nonionic surfactants are not used for several reasons, including incompatibility with the abrasive system and lack of high foaming capability.

Binding and Thickening Agent. The rheological properties of a dentifrice are primarily determined by the agent used to bind and thicken the product and allow its extrusion as a firm, but easily dispersible, ribbon. A well-formulated toothpaste exhibits a high yield point and thixotropy, that is, it is easily liquified (see RHEOLOGICAL MEASUREMENTS). Gums and resins are employed to obtain the desired thickening and binding. Each has a characteristic rheological spectrum and lends structure to the toothpaste accordingly. Gums and resins widely used include acrylic acid polymers, carrageenan [9000-07-1], sodium carboxymethyl cellulose [9004-32-4], xanthan gum [11138-66-2], and hydrated silica [10279-57-9]. Each is available in several variations having different properties. Selection of an optimal gum or resin along with the selection of appropriate other ingredients results in a paste that extrudes with the application of minimal pressure to form a smooth, cohesive ribbon, which stands up on the toothbrush bristles and breaks down and disperses quickly in the mouth during brushing.

Color. Colorants used in dentifrices are regulated by the Food and Drug Administration. They are FD&C (food, drug, and cosmetic) or D&C (drug and cosmetic) (see COLORANTS FOR FOOD, DRUGS, COSMETICS, AND MEDICAL DEVICES).

Special Active Agents. Toothpastes are vehicles for agents that provide special therapeutic and nontherapeutic effects. The scope of these agents has increased dramatically since the 1960s and research continues.

Fluoride added to a compatible dentifrice base at a level of 1000 ppm has been clinically proven to reduce the incidence of dental caries by about 25% on average, even in areas where the water supply is fluoridated (4). Elevation to 1500 ppm increases the protection. Sources of fluoride approved for use in dentifrices are sodium fluoride [7681-49-4] (0.22%), sodium monofluorophosphate (0.76%), and stannous fluoride [7783-47-3] (0.41%). The Food and Drug Administration regulates fluoridated dentifrices as drugs and has established parameters for safe and effective products. Compatibility of the fluoride with the abrasive is an important requirement.

Several agents delivered via toothpaste inhibit the accumulation of dental calculus. Pyrophosphate salts, with or without a methoxyethylene–maleic acid copolymer, and zinc salts have given positive results in clinical trials (5). Pyrophosphates were added as potassium or sodium pyrophosphate or mixtures at a level of about 2–6%. The zinc salt was zinc citrate [546-46-3] (0.5–2.0%) or zinc chloride [7646-85-7] (2.0%). The products all contained fluoride in addition to the calculus inhibitor. The anticaries activity of the fluoride was not compromised (6).

Dentifrices are also vehicles for agents that alleviate dentinal hypersensitivity. Among the materials that have given positive results in clinical tests are potassium nitrate [7757-79-1] (5%) and strontium chloride [10476-85-4] (10%).

An antimicrobial agent that reduces dental plaque and can be delivered effectively from toothpaste is a combination of Triclosan [3380-34-5] (0.2% in the toothpaste) and zinc citrate (0.5%) (7). This agent influenced plaque accumulation and reduced the incidence of gingival bleeding in clinical tests. Additional dentifrices for improved gingival health are in the offing.

Claims for safety and efficacy of therapeutic dentifrices are regulated by the Food and Drug Administration. The Council on Dental Therapeutics of the American Dental Association reviews products and grants a Seal of Acceptance to those products deemed worthy.

Specific Toothpaste Formulations. Two types of toothpaste formulation predominate. Type 1 is a low abrasive—high solvent toothpaste (Table 1); type 2 is a high abrasive—low solvent toothpaste (Table 2). The most important differences are the ratio of humectant to abrasive and the nature of the abrasive. Type 1 dentifrices were introduced nationally to the U.S. market in 1970 and now constitute the predominant type. Type 2 dentifrices represent a popular earlier formulation, in which economic and scientific considerations related to the abrasive and humectant favored use of a maximum amount of the abrasive component. All type 1 dentifrices of the early 1990s contain an amorphous hydrated silica powder as the abrasive. Type 2 dentifrices may contain one or more of many insoluble minerals.

Table 1. Type 1 Dentifrices^a

Function	Ingredient	CAS Registry Number	Amount, %
abrasive	silica xerogel (hydrated silica)	[10279-57-9]	14.00
humectant system	sorbitol, 70%	[50-70-4]	46.72
	glycerol, 96%	[56-81-5]	20.90
	poly(ethylene glycol) 1450	[25322-68-3]	5.00
surfactant	sodium lauryl sulfate	[151-21-3]	1.50
binder and thickener	carboxymethylcellulose, grade 9MX	[9004-32-4]	0.30
	silica aerogel	[7631-86-9]	8.00
flavor	flavor oils		2.00
	sodium saccharin		0.20
preservative	sodium benzoate	[532-32-1]	0.10
anticaries agent	sodium monofluorophosphate	[10163-15-2]	0.78
color solution	FD&C, D&C dyes in water		0.50

^a Ref. 8.

Table 2. Type 2 Dentifrices^a

Function	Ingredient	CAS Registry Number	Amount, %
abrasive	α-alumina trihydrate	[12252-70-9]	55.0
humectant system	glycerol	[56-81-5]	20.0
	water	[7732-18-5]	21.6
surfactant	sodium lauryl sulfate	[151-21-3]	1.5
binder and thickener	sodium carboxymethylcellulose	[9004-32-4]	1.0
flavor	flavor oils		0.9

^a Ref. 9.

Gels. Amorphous hydrated silicas of a purity and structure typical of those used in type 1 dentifrices and the liquid portion (humectant system) of type 1 dentifrices both have approximately the same refractive index, ie, about 1.47. As a result, the type 1 dentifrices represented in Table 1 are inherently transparent or translucent. In the marketplace it has become popular to refer to such dentifrices as gels. For marketing reasons some companies have chosen to opacify these products, with titanium dioxide, for example. The opacified products are identical in functionality, structure, and all other ways, except opacity, to their translucent or transparent counterparts.

Dental Rinses

Dental rinses, as discussed here, do not include gargles and other liquids that are indicated for inflammation and diseases of the throat and are unrelated to dental plaque.

The primary function of a dental rinse is to clean and refresh the mouth. With few exceptions dental rinses marketed in the early 1990s also fulfill important secondary functions. Some contain agents that inhibit or destroy oral microbial populations that aid in generating dental plaque or oral malodor. Other rinses deliver fluoride to the teeth to prevent the development of caries. A relatively new rinse is used before toothbrushing to enhance the efficacy of brushing. Dental rinses can also deliver active agents that cannot be provided by toothpaste because of chemical incompatibility between the agent and the toothpaste ingredients. For example, sodium fluoride and calcium-containing abrasives are incompatible, as are sodium lauryl sulfate (an anionic detergent) and chlorhexidine [55-56-1] (an active cationic antimicrobial agent).

The amount of a product introduced into the oral cavity is substantially greater with a dental rinse than with a dentifrice, the action of swishing the rinse assures thorough distribution of the agent to accessible oral surfaces.

An oral dental rinse generally consists of water, alcohol, a humectant, an emulsifier, flavor, color, and an active agent. Water is the primary vehicle. The alcohol provides bite and is also a formulation aid. The humectant improves the feel in the mouth and also prevents locking of the cap to the container between uses; glycerin or noncrystallizing sorbitol may be satisfactory. The emulsifier is a nonionic type, for example, a polyoxyethylene–polyoxypropylene block copolymer or a polyoxyethylene sorbitan fatty acid ester. Flavors are generally a type of mint or cinnamon. Colors are FD&C or D&C.

Active agents vary according to use. For controlling bad breath, zinc salts, sodium lauryl sulfate, and flavors are used. To destroy oral microorganisms, chlorhexidine, cetylpyridinium chloride [123-03-5], and benzalkonium chloride [68391-01-5] are valuable. Essential oils, such as thymol [89-83-8], eucalyptol [470-82-6], menthol, and methyl salicylate [119-36-8] reduce plaque-related gingivitis (see OILS, ESSENTIAL). Sodium fluoride aids in caries control.

Chlorhexidine is the most potent oral antimicrobial agent available. It has side effects and is sold only with a prescription. It is active delivered from a rinse, but a compatible toothpaste vehicle for chlorhexidine has yet to be developed.

A dental rinse containing sodium lauryl sulfate, marketed to be used before toothbrushing, removes some plaque directly and makes residual plaque easier to remove by brushing.

Claims for oral dental rinses are regulated by the Food and Drug Administration whether they are marketed as drugs or cosmetics. The Council on Dental Therapeutics of the American Dental Association reviews oral rinses, and may authorize use of the Seal of Acceptance for a product.

Economic Aspects

The dentifrice market in the United States was valued at about \$1.5 billion at the retail level in 1991. The oral rinse market has a value of about \$850 million.

Usage patterns for dentifrices vary considerably from country to country. A large portion of the market is concentrated in the industrialized countries; however, even the most undeveloped countries have markets for dentifrices, generally centered on the larger urban areas. The scale of toothpaste manufacture is variable and is directly related to market size and marketing practice. Toothpaste can be made by batch, continuously, or in combinations. Plants with outputs of 110 t or more daily have been operated, but are an exception. Toothpastes vary considerably in composition around the world, and any single enterprise may market several brands with different properties.

The strongly research-oriented dentifrice market is dominated by a few primary companies. The industry funds efforts to develop products with superior cosmetic and therapeutic performance. The four principal manufacturers of toothpaste worldwide are Procter & Gamble, Unilever, Colgate–Palmolive, and Beecham. Numerous smaller companies and private-label houses also exist.

The incidence of dental caries has decreased dramatically in recent years. It has fallen to such an extent as to reduce the need for professional dental health services related to caries significantly. The cause is not clear, but water fluoridation, addition of fluoride to toothpaste, and other modes of fluoride administration are generally conceded to be important contributors to the phenomenon and the American Dental Association recommends use of a fluoride toothpaste for all patients (10).

Toothpaste usage patterns are highly variable, even within reasonably homogeneous populations. In the United States, for example, studies have shown that about 50° of people who brush their teeth brush twice a day, 10–25° brush more than twice a day, and somewhat less than 30° brush less than twice a day (11). On average, Americans use roughly 0.9 kg of toothpaste per brusher per year. There is also tremendous variation in the effect of toothpaste because of the huge variability in motivation, brushing time, brushing proficiency, amount of toothpaste used per application, toothbrush condition, and other factors.

Tooth Whitening Agents

Tooth whitening preparations are available by dentist's prescription and over the counter. Many products are marketed as systems comprising toothpastes and treatment pastes or gels. The whitening agent in most is 10° carbamide peroxide [35220-04-3], which reacts with water to release hydrogen peroxide. The hydrogen peroxide liberates free-oxygen radicals, which bleach the dental enamel. Some preparations contain calcium peroxide [1305-79-9] as the bleaching agent (see BLEACHING AGENTS). The products have undisputed efficacy. Some dental authorities have questioned their safety, hypothesizing that long-term chronic exposure to hydrogen peroxide can result in ill effects. However, human clinical trials to substantiate this hypothesis have not yet been forthcoming, and appropriately formulated and tested whitening products must be considered safe when used in accordance with directions (12,13).

BIBLIOGRAPHY

"Dentifrices" in *ECT* 1st ed., Vol. 4, pp. 928–935, by E. Di Cyan, Di Cyan & Brown; in *ECT* 2nd ed., Vol. 6, pp. 848–852, by A. F. Posteraro; in *ECT* 3rd ed., Vol. 7, pp. 522–526, by A. F. Posteraro, College of Dentistry, New York University.

1. I. D. Mandel, *J. Clin. Periodontol.* **15**, 488 (1988).
2. P. C. Kitchin and H. B. G. Robinson, *J. Dent. Res.* **27**, 501 (1948).
3. J. J. Hefferen, *J. Dent. Res.* **55**, 563 (1976).
4. American Dental Association, *Accepted Dental Therapeutics*, 40th ed., 1984, pp. 409–412.
5. M. Pader, *Cosmet. Toiletries* **107**, 63 (1992).
6. R. R. Scruggs, P. W. Stewart, M. S. Samuels, and J. W. Stamm, *Clin. Prev. Dent.* **13**, 23 (1991).
7. C. A. Saxton and F. J. G. Van Der Ouderaa, *J. Periodontal Res.* **24**, 75 (1989).
8. M. Pader, *Cosmet. Toiletries* **102**, 81 (1987).
9. S. D. Gershon and M. Pader in M. S. Balsam and E. Sagarin, eds., *Cosmetics Science and Technology*, Vol. 14, Wiley-Interscience, New York, 1972, p. 423.
10. American Dental Association, "Guide to Dental Health," *J. Am. Dent. Assoc.* **116**, 13-G special issue (1988).

11. M. Pader, *Oral Hygiene Products and Practice*, Marcel Dekker, New York, 1988, pp. 172–173.
12. Ref. 4, p. 322.
13. M. Pader, *Chemistry and Technology of the Cosmetics and Toiletries Industry*, Blackie Academic & Professional, London, 1992, pp. 253–254.

General References

American Dental Association, "Guide to Dental Health," *J. Am. Dent. Assoc.* **116**, special issue (1988).
G. Nikiforuk, *Understanding Dental Caries*, S. Karger, Basel, 1985.
M. Pader, *Oral Hygiene Products and Practice*, Marcel Dekker, New York, 1988.

Morton Pader
Consumer Products Development Resources, Inc.



DERMATOLOGICALS.

See ANTIAGING AGENTS; COSMETICS.

DESICCANTS

Many substances take up (sorb) water from their surroundings by one or more of a number of different physical or chemical mechanisms. Many common materials have an affinity for water, such as wood, paper, natural fibers, polymers, solvents, or salts. Some materials have sufficient capacity for water or efficiency for drying as well as appropriate physical and chemical properties that they are classed as drying agents, or desiccants. These substances are widely used for removing water from gases, liquids, and solids. Desiccants may be liquids or solids. They may be used repetitively by regenerating the desiccant after use to return it to its active state, or they may be used only once. If the desiccant is used only once, it may last the life of the article being dried or may be discarded when spent. Drying agents are used either in a static (batchwise) or dynamic (continuous or semicontinuous) mode. Their use may be further classified as open system, if fluid flows through the system, or closed system if it does not. Examples of the industrial uses of desiccants, designated as dynamic or static and open- or closed-system applications, are given in Table 1. The list is not all inclusive and ignores various laboratory uses.

Table 1. Applications of Desiccants

Industry	Application	Classification
compressed air	prevent freeze-up and corrosion in air-actuated components	dynamic, open
air separation	prevent ice formation in heat exchangers before cryogenic distillation	dynamic, open
natural gas	prevent corrosion and hydrate formation in pipelines, remove water before cryogenic hydrocarbon recovery, dry liquefied petroleum gas (LPG) to prevent freeze-ups during vaporization	dynamic, open
petrochemical	remove moisture before low temperature fractionation, remove moisture during the rejuvenation or burnoff of spent catalysts, prevent side reactions during catalytic refining	dynamic, open
chemical	remove water that is a diluent or contaminant of some finished product, remove water prior to or during polymerization reactions, prevent caking and corrosion	static or dynamic, closed or open
storage and shipping	prevent food deterioration and corrosion of equipment by relative humidity control	static or dynamic, closed or open
moisture vapor control	lower the dew point in sealed spaces where condensation could occur	static, closed
vapor compression refrigeration	remove moisture from circulating refrigerants	dynamic, closed
space cooling	dry air to permit cooling by evaporation of water	dynamic, open
dehumidification	dry ambient air for air-conditioning or for storage, manufacture, or drying of moisture-sensitive parts or materials	dynamic, open
corrosion control	reduce the dew point in automobile exhaust systems during cool down to reduce internal cold condensate corrosion	static, semiclosed
absorption refrigeration	cyclic absorption and stripping of water with liquid desiccants to produce chilled water for air conditioning ^a and process cooling	dynamic, open

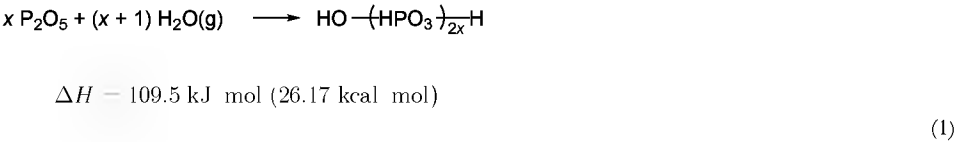
^a See AIR CONDITIONING

Desiccants have varied fundamental characteristics in terms of water capacity and the rate of water sorption. The degree of water removal achieved, or efficiency, is usually given in terms of the water content remaining in the substance that has been dried. This water content can be expressed in several ways, such as humidity ratio, relative humidity (at atmospheric pressure only), relative saturation (at elevated pressure or in liquids), dew point (used at any temperature), ice or frost point (used below 0°C), or parts per million (ppm) by weight or volume. The effectiveness of any drying agent can be measured in terms of its water capacity. In static applications, this capacity is usually the true equilibrium capacity. In dynamic systems, the rate of water removal must be taken into account. Usually, to allow for mass-transfer zones, an additional amount of drying agent is used. In these instances the dynamic capacity, also termed breakthrough capacity, falls short of the true equilibrium capacity (1).

Mechanism

The drying mechanisms of desiccants may be classified as follows: Class 1: chemical reaction, which forms either a new compound or a hydrate; Class 2: physical absorption with constant relative humidity or vapor pressure (solid + water + saturated solution); Class 3: physical absorption with variable relative humidity or vapor pressure (solid or liquid + water + diluted solution); and Class 4: physical adsorption.

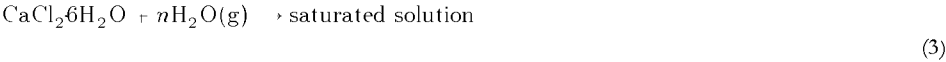
These mechanisms are characterized by the relative magnitudes of the heats of reaction, solution, or adsorption (see ADSORPTION, SEPARATION). All useful drying mechanisms are exothermic. Phosphorus pentoxide is a Class 1 drying agent that reacts with water to form a polyphosphoric acid (2):



Class 1 drying agents (and zeolites, which are Class 4 desiccants) liberate the largest amounts of heat and should be used with appropriate care. Calcium chloride is a Class 1 drying agent that reacts with water to form a hydrate:



The more highly complexed hydrates of calcium chloride ($\text{CaCl}_2\cdot n\text{H}_2\text{O}$ where $n > 2$) may also exhibit the characteristics of a Class 2 drying agent, because the hydrated species can physically absorb additional water to form a saturated solution. The term absorption is used to describe the phenomenon that occurs when a gas or vapor penetrates the solid structure to produce a saturated solution:



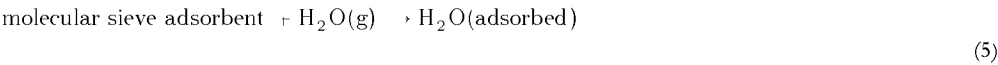
Because calcium chloride has a number of hydrates, the one that is in equilibrium with a saturated solution is a function of the temperature. In this case, the solid is dissolved as it absorbs water to form the saturated solution, and three phases are present: solid, saturated solution, and vapor. Systems having these three phases, or two solids and a vapor phase, have a constant vapor pressure at a given temperature. Therefore, Class 2 drying agents can be used to maintain a constant relative humidity.

Ethylene glycol is an example of a Class 3 drying agent. Because the solution produced is unsaturated only two phases, solution and vapor, exist:



For this system, the vapor pressure is a function of both temperature and the concentration of water in the dilute solution.

Molecular sieve zeolites are an example of a Class 4 drying agent (see MOLECULAR SIEVES). Water is removed by physical adsorption, but at no time does the adsorbent change phase or dissolve.



Adsorption (qv) is a phenomenon in which molecules in a fluid phase spontaneously concentrate on a solid surface without any chemical change. The adsorbed molecules are bound to the surface by weak interactions between the solid and gas, similar to condensation (van der Waals) forces. Because adsorption is a surface phenomenon, all practical adsorbents possess large surface areas relative to their mass.

Desiccants can lose water capacity and drying efficiency by taking up moisture during storage. They should therefore be analyzed before use. If necessary, the materials should be reactivated (regenerated) before putting them in service.

Compatibility

Desiccants must be chemically compatible with the material being dried. Ideally, the desiccant and the material should not react because such a reaction may produce harmful or undesirable by-products. For example, in the refrigeration and air-conditioning industry, new ozone-safe refrigerants are being proposed and tested to replace the chlorofluorocarbons in vapor compression refrigeration systems. The desiccants that keep these fluids dry must not react appreciably with the refrigerants or lubricants in the systems. Compatibility tests are conducted in which the desiccant and fluid are contacted under pressure often in the presence of system materials, such as lubricants and metals (3). The mixture is aged for various periods, typically seven days or more, at temperatures above those expected at the desiccant location in the refrigeration system to increase the reaction rates. Following the exposure, the desiccant is analyzed for degradation products and retention of water capacity and physical properties.

Static Drying

Many liquids are dried batchwise rather than continuously. The drying agent is added to the liquid and sufficient time is allowed to dry the product. The liquid is then separated from the drying agent by filtration, decantation, or distillation. Drying agents employing Class 1 or 2 mechanisms are generally used for these applications.

Desiccants Used in Static Drying. The most commonly used desiccants are discussed in this section; activated alumina, silica gel, and molecular sieves, which are discussed later under dynamic, solid drying agents, are also widely used in static or batch-drying situations.

Barium Oxide. This compound [1304-28-5] (Class 1, nonregenerative) is used primarily as a laboratory drying agent (4). It is the only drying agent that continues to dry even at red heat. Barium oxide is relatively expensive and cannot be regenerated by conventional methods. Therefore, it is not used extensively in commercial applications (see BARIUM COMPOUNDS).

Calcium Chloride. Calcium chloride, [10043-52-4], CaCl_2 , (Class 1 or 2, regenerative), can be either a solid or liquid drying agent (4). Its principal advantage is that its cost is low enough to permit discarding after use in small units. Commercial anhydrous calcium chloride is available in a range of compositions from $\text{CaCl}_2\cdot 0.05\text{H}_2\text{O}$ to $\text{CaCl}_2\cdot 0.25\text{H}_2\text{O}$. Figure 1 is a phase diagram for calcium chloride (see CALCIUM COMPOUNDS).

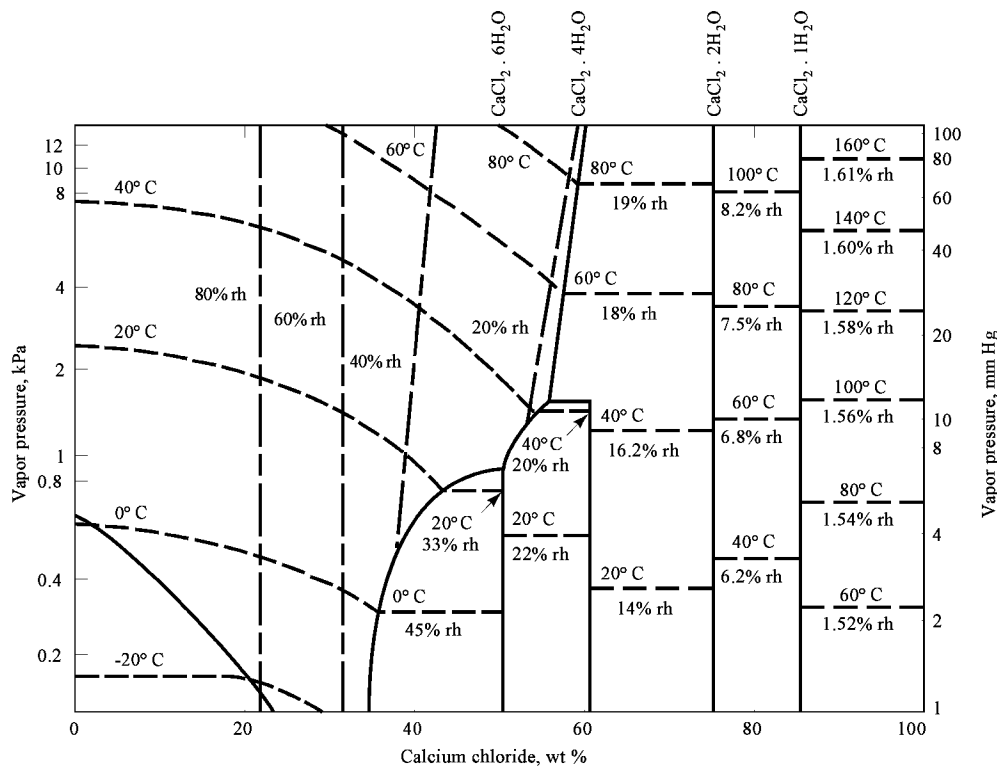


Fig. 1. Vapor pressure and relative humidity over CaCl_2 solutions and solids. The straight horizontal lines in the right-hand portion represent two solid phases and a gas phase for vertical line intersections. In addition, a solid phase, saturated solution, and a vapor phase occur in the regions between the vertical lines. The lower left-hand corner shows the ice solution line. The region in between, with skewed isothermal lines, represents unsaturated solutions; the vapor pressure varies as a function of temperature.

Calcium Oxide. Also called lime or quicklime (4,5), calcium oxide [1305-78-8] (Class 1, nonregenerative), is relatively inexpensive. It is prepared by roasting calcium carbonate (limestone) and is available in a soft and a hard form according to the way in which it was burned. For desiccant service, soft-burned lime should always be used. Calcium oxide is most commonly used to dehydrate liquids and is most efficient when it can be heated to speed the reaction rate. The reaction product is calcium hydroxide, which crumbles as it picks up moisture.

Calcium Sulfate. Calcium sulfate [7778-18-9] (Class 1, regenerative) is sold under the trade name Drierite (4,6). It occurs in nature in the anhydrous form, CaSO_4 , and in the hydrated form, $\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$, commonly known as gypsum. When prepared properly, small capillaries form within the granules and increase the somewhat low water capacity of the material by sorbing additional moisture. The first stages of water removal occur by adsorption and a chemical reaction to form a hemihydrate [10034-76-1], $\text{CaSO}_4 \cdot \frac{1}{2}\text{H}_2\text{O}$. The material can be regenerated repeatedly by heating to about 200°C. However, above 300°C, it loses some of its desiccating power. Calcium sulfate is used extensively because it is chemically inert to most materials, reusable, and inexpensive.

Lithium Chloride. Of the metal halides, calcium bromide [7789-41-5], CaBr_2 , zinc chloride [7646-85-7] ZnCl_2 , CaCl_2 , and lithium chloride [7447-41-8] LiCl , (Class 1, nonregenerative) are the most effective for water removal (4). All are available in the form of deliquescent crystals. The hydrates of LiCl are $\text{LiCl} \cdot n\text{H}_2\text{O}$, where $n = 1, 2, \text{ or } 3$. Lithium chloride solutions are more stable in air and less corrosive than the other metal halides. The high solubility of lithium carbonate [554-13-2], Li_2CO_3 , usually eliminates scale formation problems (see LITHIUM COMPOUNDS).

Perchlorates. The three common perchlorates (Class 1, nonregenerative) used as drying agents are barium perchlorate [13465-95-7], $\text{Ba}(\text{ClO}_4)_2$, lithium perchlorate [7791-03-9], LiClO_4 , and magnesium perchlorate [10034-81-8] $\text{Mg}(\text{ClO}_4)_2$. The last is the most efficient with drying action above 100°C. Even the higher hydrate form has good drying capacity. Perchlorates are strong oxidizing agents and should never be used in the presence of organic compounds because the mixture is highly explosive. For this reason, perchlorates are usually not regenerated (see CHLORINE OXYGEN ACIDS AND SALTS).

Phosphorus Pentoxide. This compound, P_2O_5 , (Class 1, nonregenerative) is made by burning phosphorus in dry air. It removes water first by adsorption, followed by the formation of several forms of phosphoric acid (2). Phosphorus pentoxide [1314-56-3] has a high vapor pressure and should only be used below 100°C. Its main drawback is that as moisture is taken up, the surface of the granules becomes wetted and further moisture removal is impeded. For this reason, phosphorus pentoxide is sometimes mixed with an inert material (see PHOSPHORIC ACIDS AND PHOSPHATES).

Sodium and Potassium Hydroxides. Sodium hydroxide [1310-73-2] and potassium hydroxide [1310-58-3] (Class 1, nonregenerative) are commonly used when moisture and carbon dioxide or hydrogen sulfide must be removed simultaneously (4). Fused sticks or solutions of the alkali hydroxides are frequently used. These materials must be handled with care to prevent serious skin burns.

Capacity and Efficiency. Figure 2 shows the drying capacity of selected desiccants as a function of relative humidity. The higher capacity desiccants go through the various hydrate levels to yield fairly broad ranges of constant relative humidities as moisture is picked up. However, these compounds do not produce very low relative humidities or dew points. Figure 3 shows the water-vapor pressure over several desiccants as a function of temperature. The addition of more desiccant lowers the vapor pressure. The best performance, or lowest dew point, occurs with excess drying agent. The minimum dew points attainable in air at room temperature and atmospheric pressure are given in Table 2.

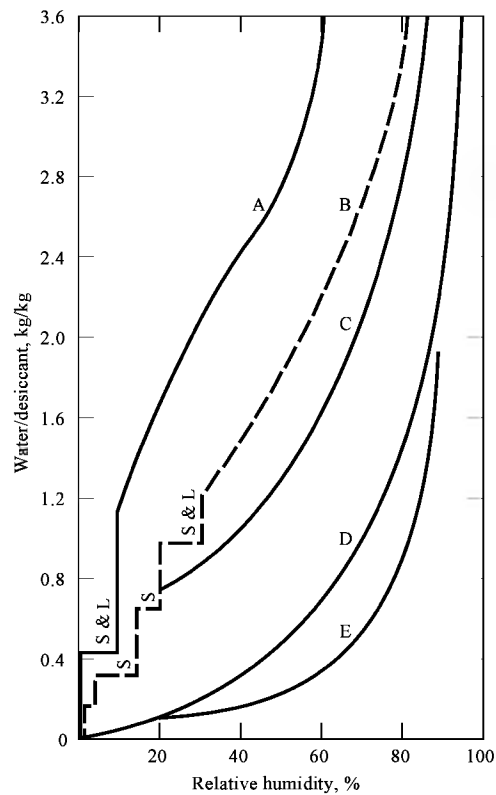


Fig. 2. Drying capacity of selected drying agents in liquid form. A represents lithium chloride; B, calcium chloride; C, sulfuric acid; D, glycerol; E, triethylene glycol. S is solid, L is liquid.

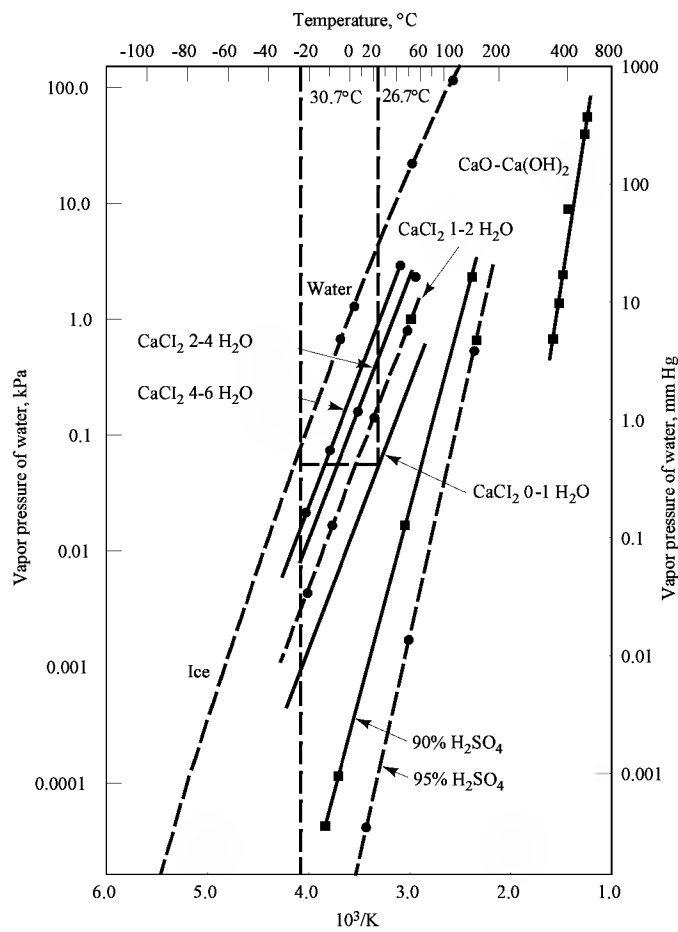


Fig. 3. Water-vapor pressure over ice and several drying agents.

Table 2. Performance of Chemical Desiccants in Drying Air at Room Temperature

Substance	Minimum humidity ^a	Minimum dew point at 101.3 kPa, ^b °C
P ₂ O ₅	0.0193	98
BaO	0.503	80.5
KOH fused	1.546	73

CaO	2.32	−70.5
H ₂ SO ₄	2.32	−70.5
CaSO ₄ anhyd.	3.87	67
Al ₂ O ₃	3.87	67
KOH sticks	10.83	−60
NaOH fused	123.7	−40
CaBr ₂	139.2	39
CaCl ₂ fused	262.9	33
NaOH sticks	618.6	25
Ba(ClO ₄) ₂	634.0	24.5
ZnCl ₂	657.2	24
ZnBr ₂	896.9	21
CaCl ₂ granular	1159.8	18

^a mg H₂O / kg dry air.
^b To convert kPa to mm Hg, multiply by 7.5.

The efficiency ranking of desiccants in drying air is not always the same as that observed in drying other materials. Other materials may interact with the desiccants to reduce drying effectiveness. From a study of the efficiency of some 25 desiccants for drying several families of laboratory solvents and reagents it was concluded that molecular sieves are the desiccants of choice in most cases (9–17).

Closed-System Drying. Equilibrium capacity is the principal consideration in the design of closed nonregenerative, relatively static drying systems. The total amount of moisture to be removed must first be calculated from the volume of the system and the initial water concentration. Depending on the final moisture content desired, a drying agent can be selected based on its compatibility with the material to be dried and its ability to produce the final dew point. The equilibrium capacity of the drying agent must be determined at the system temperature and the final water concentration. An amount of drying agent must be used so that the total amount of water to be removed does not exceed the predetermined equilibrium capacity of the agent. The agent and the material are then placed in intimate contact. After sufficient time, the system comes into equilibrium with the drying agent.

Although equilibrium capacity is the prime concern, few of these closed systems are truly static. Even closed systems have dynamic features or other nonsteady-state aspects, such as temperature fluctuations, moisture ingression, and drydown rates.

Closed systems are usually nonregenerative: the desiccant charge is designed for the life of the system or is replaceable. Often in a nominally closed (static) drying system, the total amount of moisture to be removed consists of both the water initially in the system and that which leaks into the system over its lifetime. In multipane (insulating glass) windows, for example, molecular sieve desiccant is used to keep the air (or gas-filled) space dry to prevent condensation in cold weather. The glass panes are typically separated by an aluminum spacer, which is a hollow channel partially filled with the desiccant. The glass and spacer are joined and sealed with organic sealants. Although a small amount of desiccant is required to dry the gas initially sealed into the space, a much larger amount is typically used to adsorb moisture that is slowly transmitted into the space from the outside. The transmission of water vapor through the sealant is driven by the difference in water partial pressure between the desiccated space and the ambient air. Thus to specify the mass of desiccant needed requires knowledge of the moisture vapor transmission rate and specification of the design life of the unit as well as the moisture content of the gas initially sealed into the space.

Because the system likely is nonisothermal, the analysis of a closed-desiccant system requires knowledge of the temperature of the desiccant as well as the dew point (ice point) or water concentration (partial pressure) specification. Indeed, the whole system may undergo periodic temperature transients that may complicate the analysis. For example, in dual-pane windows the desiccant temperature is approximately the average of the indoor and outdoor temperatures after a night of cooling. However, after a day in the sun, the desiccant temperature becomes much warmer than the outdoor temperature. When the sun sets, the outdoor pane cools quickly while the desiccant is still quite warm. The appropriate desiccant for such an application must have sufficient water capacity and produce satisfactory dew points at the highest temperatures experienced by the desiccant.

Another aspect to consider in the design of closed-drying systems is the drydown time. The drydown time is the period required for the system to dry down from its initial water concentration (or partial pressure) to a concentration that approaches equilibrium with the desiccant. During this time, the system is not fully protected from the negative effects of the moisture that the desiccant is designed to remove. In such a system, the instantaneous drying rate is proportional to the water content at any time (18).

The drying rate is represented by differential equation (eq. 6) where *b* is mass transfer coefficient 1/(h·cm²); *A*, specific surface area of desiccant beads, cm²/g; *m*, mass of desiccant, g; *C*, concentration by weight of water in the fluid being dried; *C*_s, concentration of water at the surface of the desiccant, ie, concentration of water in the fluid that would be in equilibrium with the instantaneous loading on the desiccant, wtppm; and *t* = time, h.

$$\frac{dC}{dt} = -bAm(C - C_s)$$

(6)

If *C*_s is approximately constant during the initial drydown period, as it is in many closed-system applications, then the water concentration decays exponentially with time. The rate equation can be integrated to give the relationship between water concentration and time:

$$\frac{C}{C_0} = \exp[-bAM(t - t_0)]$$

(7)

where *C*₀ is the water concentration at the initial time *t*₀.

If required, the drydown can be hastened by increasing desiccant mass, particle surface area, or mass-transfer coefficient. The mass-transfer coefficient can be altered to some extent by the design of the desiccant container.

Because a material balance on water must be satisfied during the drydown as well as afterward, the path from the initial concentration to equilibrium can be represented graphically by a material balance line and an equilibrium curve. The coordinates of the starting point on the material balance line are the initial water contents of the fluid to be dried and the desiccant. The slope of the line is the ratio of fluid mass to desiccant mass. The line terminates at its intersection with the equilibrium curve (Fig. 4).

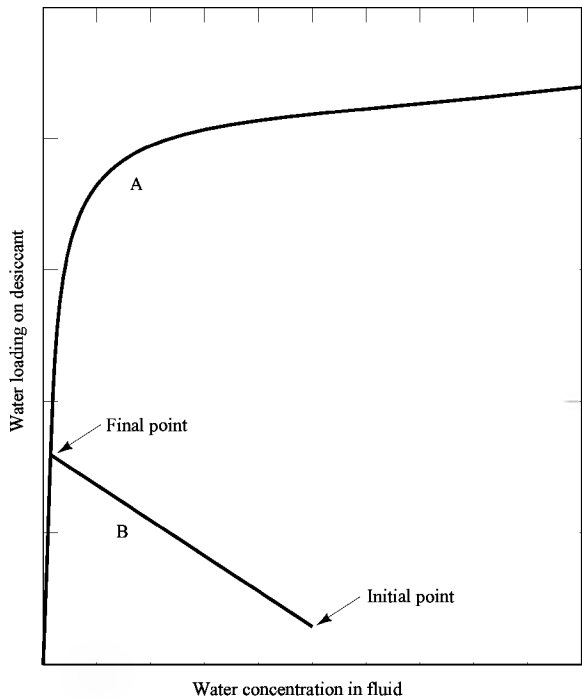


Fig. 4. Drydown path to equilibrium in a closed system. A represents the equilibrium curve and B, the material balance line.

An interesting and novel use of a solid desiccant, the reduction of cold condensate corrosion in automotive exhaust systems, illustrates a hybrid closed–open system. Internal corrosion occurs in mufflers when the water vapor in the exhaust condenses after the engine is turned off and the muffler cools. Carbon dioxide dissolves in the condensate to form an acidic soup. In an essentially closed static drying step, an acid- and heat-resistant desiccant located in the muffler adsorbs water vapor from the exhaust gas as it cools to prevent formation of corrosive acidic condensate. When the engine is restarted, the system becomes open, and the desiccant is regenerated by the hot exhaust gas to be ready for the next cooldown step (19).

Dynamic Desiccants

Continuous drying is employed when drying a volume of gas or liquid in a batchwise fashion is not practical. When a solid, dynamic desiccant is used, the fluid stream is passed over a fixed bed of the drying agent, which must have the physical properties to allow the fluid to pass readily through. When liquid desiccants are used, the drying is usually achieved by countercurrent contact of the gas (flowing up) against the liquid (flowing down). The desiccants are usually regenerable.

Liquid Desiccants. Glycols and sulfuric acid are the principal examples.

Sulfuric Acid. H_2SO_4 [7664-93-9] (Class 3, regenerative) is used extensively throughout the chemical industry to dry acidic and corrosive gases.

It has good capacity and drying capability as illustrated by the vapor pressure curves in Figure 5. At 25°C, the dew point attainable in gases dried with 95% sulfuric acid is less than 75°C.

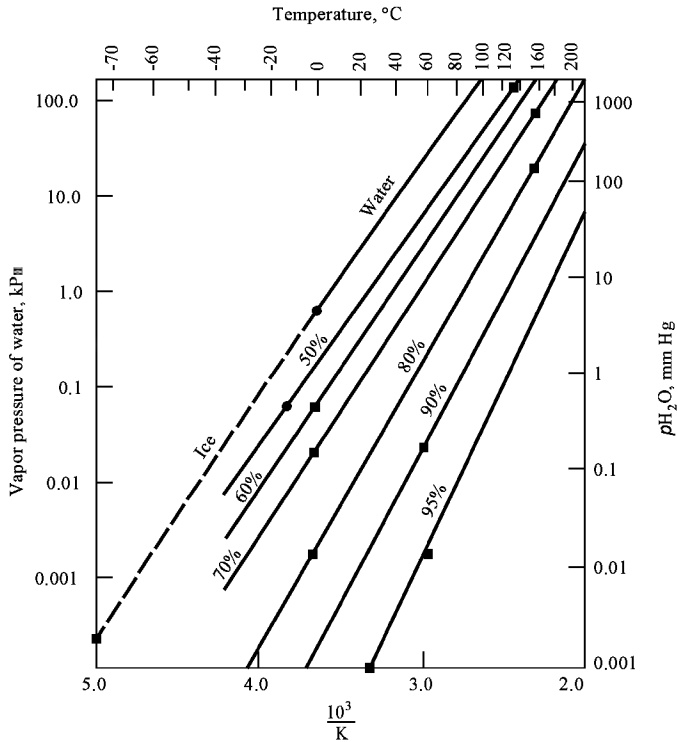


Fig. 5. Vapor pressure of water over sulfuric acid solutions. Percentage of H_2SO_4 noted on each curve.

Sulfuric acid is used in circulating towers like those depicted in Figure 6; the gas flows countercurrent to the acid. The spent acid is removed from the primary contactor at a 50% H_2SO_4 content and may be used for chemical manufacturing or recycled after reconcentration. The makeup, or recycled, acid is usually introduced at 93% H_2SO_4 (sp gr 1.84, 66° B $\ddot{\text{u}}$ or higher) because this moisture content establishes the final moisture level of the dried product gas. Sulfuric acid is highly corrosive, and protective clothing and eye protection must be provided.

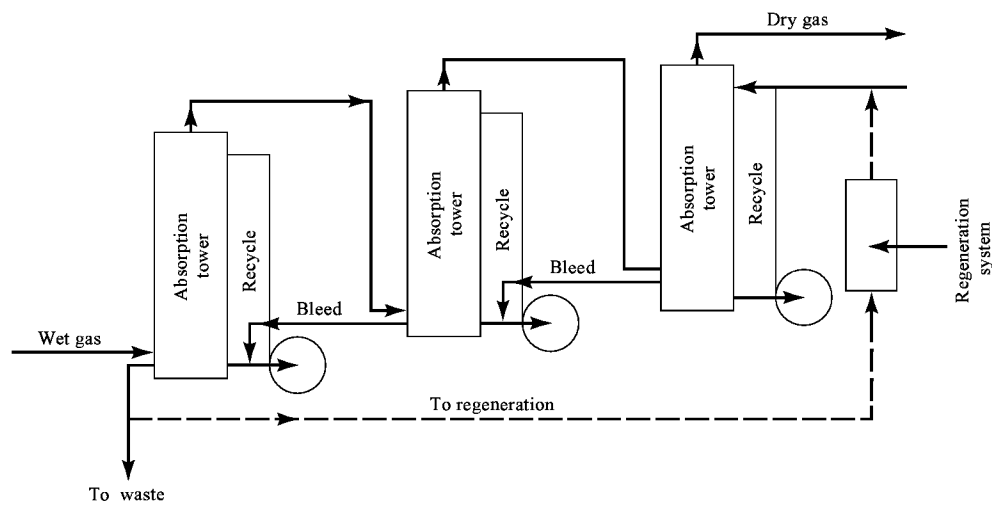


Fig. 6. Continuous sulfuric acid drying system.

Glycerol, Glycol, and Other Polyhydric Alcohols. These alcohols (Class 3, regenerative) are widely used to dry gases (20,21). However, they can only produce dew points in the range of -15 to 0°C (see ANTIFREEZES AND DEICING FLUIDS). They have somewhat lower capacity than sulfuric acid (Fig. 2) but are effective when either injected or employed in a multistage contactor to achieve dew point depression.

Ethylene glycol [107-21-1], diethylene glycol [111-46-6], and triethylene glycol [112-27-6] are used extensively in the natural gas industry to inhibit hydrate formation (22,23). Diethylene glycol (DEG) is the most widely used. The dew point that can be achieved in the treated gas (24,25) is a function of the percentage of glycol present and the contact temperature (Fig. 7). The principal advantages of glycol dehydration are low cost, ease of regeneration, and minimal losses of the drying agent caused by solubility or vapor pressure during subsequent recovery or reclamation.

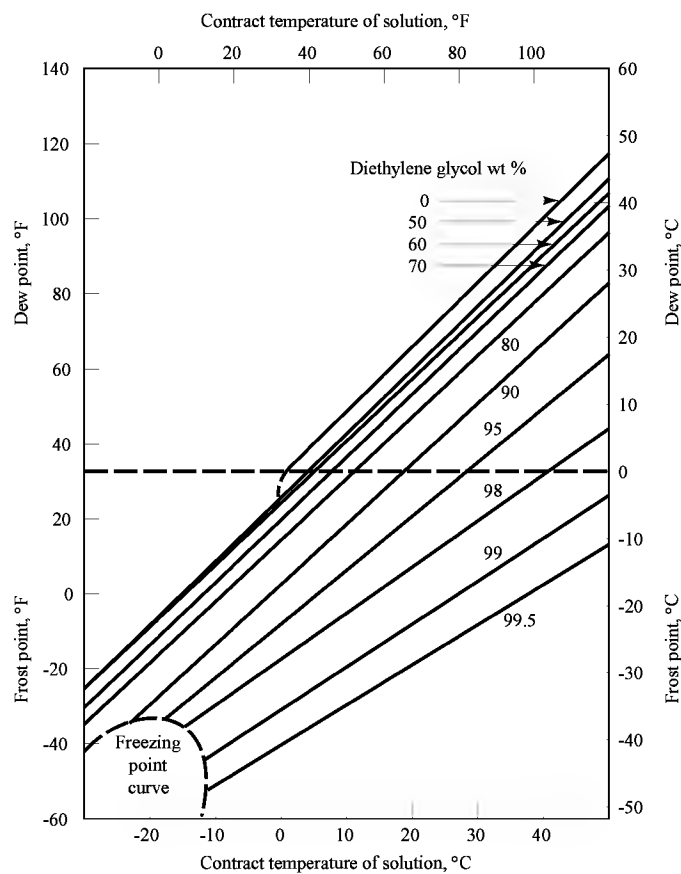


Fig. 7. Dew points of aqueous diethylene glycol solutions.

A typical flow diagram of a glycol dehydration plant is shown in Figure 8. The absorber, or gas-liquid contactor, operates with glycol flowing downward, countercurrent to the gas stream. The regenerator, or stripping column, operates at low pressures to keep the column temperature below the decomposition temperature of the glycol. The regenerated glycol is then recirculated to the absorber (see ALCOHOLS, POLYHYDRIC; GLYCEROL; GLYCOLS).

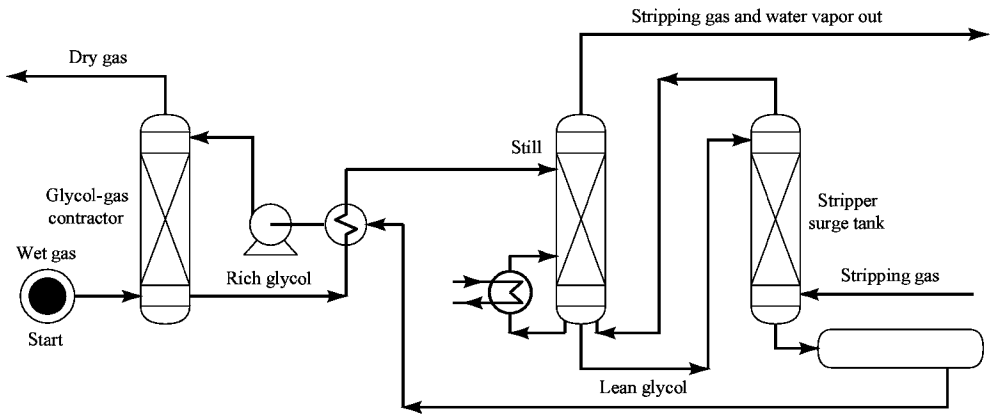


Fig. 8. Natural gas diethylene glycol dehydration system.

Solid Desiccants. The solid desiccants used in dynamic applications fall into a class called adsorbents (see ADSORPTION). Because they are used in large packed beds through which the gas or liquid to be treated is passed, the adsorbents are formed into solid shapes that allow them to withstand the static (fluid plus solid head) and dynamic (pressure drop) forces imposed on them. The most common shapes are granules, extruded pellets, and beads.

Activated Alumina. This material (Class 4, regenerative) is made by the calcination of an alumina gel or aluminum oxide trihydrate [21645-51-2], $\text{Al}_2\text{O}_3 \cdot 3\text{H}_2\text{O}$, into various crystalline phases of transition aluminas (26). Depending on the manufacturing procedure and starting material, the final product has different degrees of specific surface and pore volume. Table 3 lists the physical properties of a typical Grade A activated alumina.

Table 3. Physical Properties of Solid Adsorbents

Property	Grade A activated alumina	Silica gel	Type 4A molecular sieve
surface area, m ² g	320	832	750
bulk density, kg m ³	800	720	670
maximum heat of adsorption, J g H ₂ O ^a	ca 1395	ca 930	ca 4180
specific heat, J (kg·K) ^a	1005	921	1046
reactivation temperature, °C	150–315	125–275	200–315
pore volume, % of total	ca 50	ca 55	ca 48
pore size, nm	1–7.5	1–40	0.42
pore volume, cm ³ 100 g	40	43	28.9

^aTo convert J to cal, divide by 4.184.

The water removal mechanism is adsorption, which is the mechanism for all Class 4 drying agents. The capacity of such materials is often shown in the form of adsorption isotherms as depicted in Figures 9a and 9b. The initial adsorption mechanism at low concentrations of water is believed to occur by monolayer coverage of water on the adsorption sites. As more water is adsorbed, successive layers are added until condensation or capillary action takes place at water saturation levels greater than about 70% relative humidity. At saturation, all the pores are filled; and the total amount of water adsorbed, expressed as a liquid, represents the pore volume of the adsorbent.

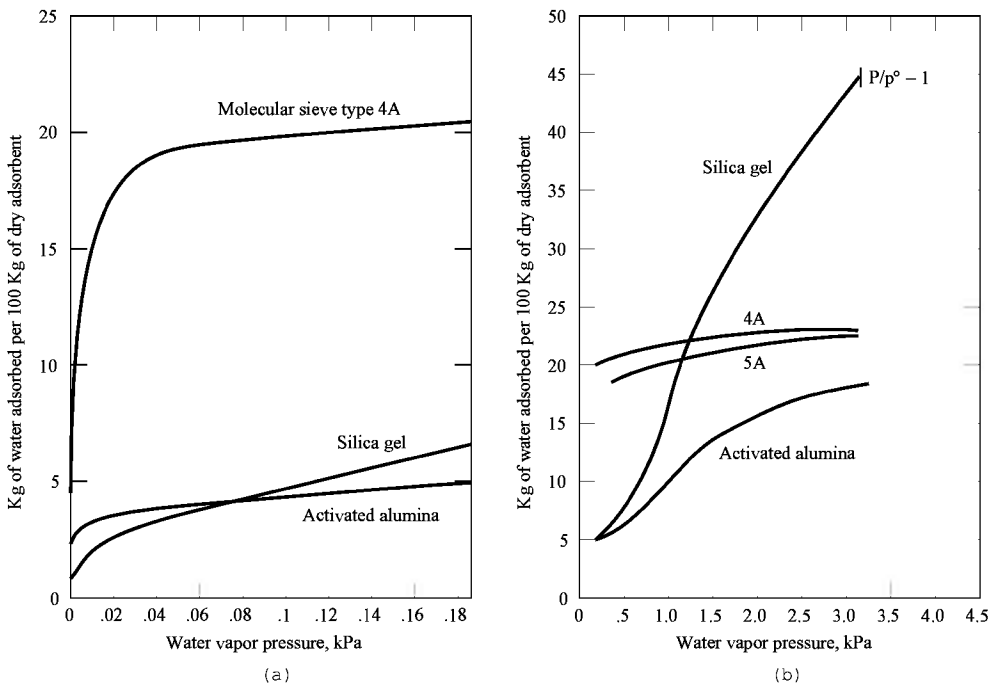


Fig. 9. Water adsorption isotherms at 25°C (to convert kPa to mm Hg, multiply by 7.5). 4A and 5A designate molecular sieves.

In regenerating activated alumina, heating to a temperature of 150 to 200°C is sufficient to recover nearly all of the initial water capacity. This regeneration is usually accomplished by passing a heated gas through the adsorbent bed. Unless the surface is fouled by the gas or liquid being dried, the life of activated alumina is good and usually depends on the number of regeneration cycles (see ALUMINUM COMPOUNDS).

Silica Gel. This material [7631-86-9] (Class 4, regenerative) is made by dehydration of high purity silica hydrogel (27). The final product is high in purity (99.7% SiO₂), which contributes to its chemical inertness. The typical physical properties are listed in Table 3. The pore size of silica gel has the broadest range of the three primary solid desiccants. Therefore, it can adsorb larger molecules in addition to water (critical diameter = 0.265 nm). The average pore diameter is about 2.5 nm.

The capacity of silica gel is shown in Figures 9a and 9b, and the shape of the isotherm is similar to activated alumina. At saturation ($p/p^{\circ} = 1.0$), silica gel, which takes up 40 kg H₂O / 100 kg of adsorbent, has the highest capacity of the desiccants shown. However, some high capacity silica gels tend to shatter in the presence of liquid water. When liquid water may be present, a lower capacity, water-resistant silica gel must be used.

The normal regeneration temperature for silica gel is 175°C. In hydrocarbon service, higher temperatures (225–275°C) are recommended to desorb heavy hydrocarbons, which tend to foul the adsorbent during prolonged use (see SILICON COMPOUNDS).

Molecular Sieves. Molecular sieve desiccants (Class 4, regenerative) are members of a class of materials called zeolites (28). Although some occur in nature, commercial molecular sieve zeolites are usually synthetic (see MOLECULAR SIEVES). They are crystalline framework aluminosilicates containing alkali metal cations. The structure extends in three dimensions by a network of AlO₄ and SiO₄ tetrahedra linked to one another by the sharing of oxygen atoms. Molecular sieves possess the high porosity that is characteristic of all adsorbents. In addition, the ordered crystalline structure of the molecular sieve provides pores of a constant size. In contrast, the pores in activated alumina and silica gel are nonuniform in size, as shown in Table 3. Silica gel is an amorphous material with no crystal structure. The pore structure in activated alumina is not contained within the alumina crystals but is formed by the spaces between randomly agglomerated crystals.

The pore size of molecular sieves can be enlarged or diminished by appropriate cation exchange. Many commercial types are available with pore openings ranging from 0.3 nm to about 1.0 nm. For example, type 3A (the potassium form of zeolite A) has a nominal pore opening of 0.3 nm (3 Å) and type 4A (sodium form) has a nominal 0.4 nm opening. Type 5A (calcium form) with a 0.5 nm opening and type 13X (sodium form of zeolite X) with an 0.85 nm opening are also available.

All these forms adsorb water molecules. The constant size and adjustable pore opening permit the exclusion of many other gaseous and liquid molecules from the internal pore structure, hence the name molecular sieve. This feature provides unique advantages in certain applications. Fouling of the adsorption surface by compounds with high molecular weights can be prevented by excluding them. Molecular sieves with larger pores can also be used simultaneously to dry and purify process streams, eg, by the adsorption of carbon dioxide or sulfur compounds in addition to water.

Because of their ordered structure, molecular sieves have high capacity at low water concentrations and do not exhibit a capillary condensation pore-filling mechanism at high water concentrations. The desiccating properties of the material are still good at elevated temperatures (Fig. 10). A dew point of -75°C can be obtained in a gas dried at 90°C with a molecular sieve that adsorbs water to the level of 1 wt %. In normal operations at ambient temperature, dew points of < -100°C have been measured.

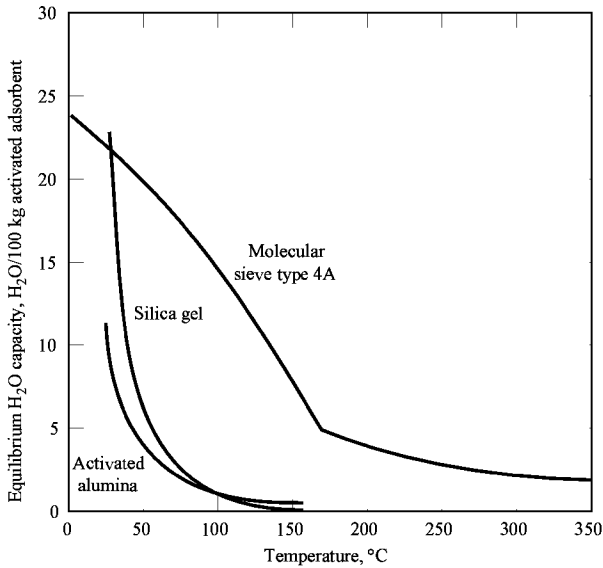


Fig. 10. Adsorbent isobars ($p_{\text{H}_2\text{O}} = 1.38 \text{ kPa}$). To convert kPa to atm, divide by 101.3.

Molecular sieves are also inert to most fluids and are physically stable when wetted with water. Strong inorganic acids or alkalies and temperatures above 700°C should be avoided. Mildly acidic streams can be dried with a molecular sieve of an acid-resistant type (29).

Design of Dynamic Adsorption Drying Systems

Adsorbent drying systems are typically operated in a regenerative mode with an adsorption half-cycle to remove water from the process stream and a desorption half-cycle to remove water from the adsorbent and to prepare it for another adsorption half-cycle (8,30,31). Usually, two beds are employed to allow for continuous processing. In most cases, some residual water remains on the adsorbent after the desorption half-cycle because complete removal is not economically practical. The difference between the amount of water removed during the adsorption and desorption half-cycle is termed the differential loading, which is the working capacity available for dehydration.

The two most common types of drying systems operate on either a pressure-swing cycle or a thermal-swing cycle to take advantage of the difference in water loading on the desiccant with changes in pressure and temperature (Fig. 11) (28). A pressure-swing cycle uses a high pressure adsorption step and a low pressure desorption step and does not require an elevated temperature for regeneration. However, some amount of thermal energy can be added, if desired, during the desorption half-cycle to increase the desorption efficiency. This type of system operates with small differential water loadings. Therefore, short adsorption and desorption times are used (1–60 min). Higher effluent dew points are characteristic of this type of dehydration operation. A thermal-swing cycle requires an elevated temperature during the desorption step. Depressurization can also be used during this step to improve regeneration efficiency. This type of cycle operates with the highest differential loadings, and longer cycle times are normally used (4–24 h). If low dew points are required, a thermal-swing cycle should be employed.

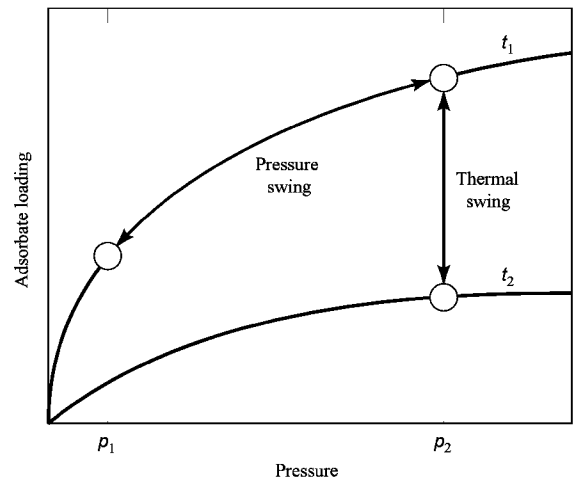


Fig. 11. Schematic illustration of the thermal-swing and pressure-swing cycles. In thermal swing, the differential loading, Δx , is given by: $\Delta x_t = x_{t1} - x_{t2}$ at p_2 ; in pressure swing: $\Delta x_p = x_{p2} - x_{p1}$ at t_1 .

In many cases a two-bed dehydrator system, with one bed on adsorption and the other on regeneration, can process all the fluid to be dried. Figure 12 is a schematic of a simple two-bed natural gas dehydrator; the same scheme can be used to dry air. Natural gas is first passed through a gas-liquid separator to ensure single-phase operation. The gas passes downflow through a molecular sieve drying tower. The dried product natural gas may then be sent for further processing, eg, cryogenic hydrocarbon recovery of liquefied petroleum gas (LPG) condensates. The residue or lean gas from the cold section of such a plant is commonly used for regeneration. The gas is heated and passed through the exhausted adsorbent bed (Tower B) with the flow countercurrent to the gas drying step. Tower B is then cooled to feed temperature by flowing gas around the heater directly to the bed. The spent regeneration gas then passes through a cooler, and liquid water is removed before it is returned to the pipeline. Depending on the process conditions, the amount of regeneration gas required is about 5–10% of the total amount of gas dried (see CRYOGENICS; GAS, NATURAL; LIQUEFIED PETROLEUM GAS).

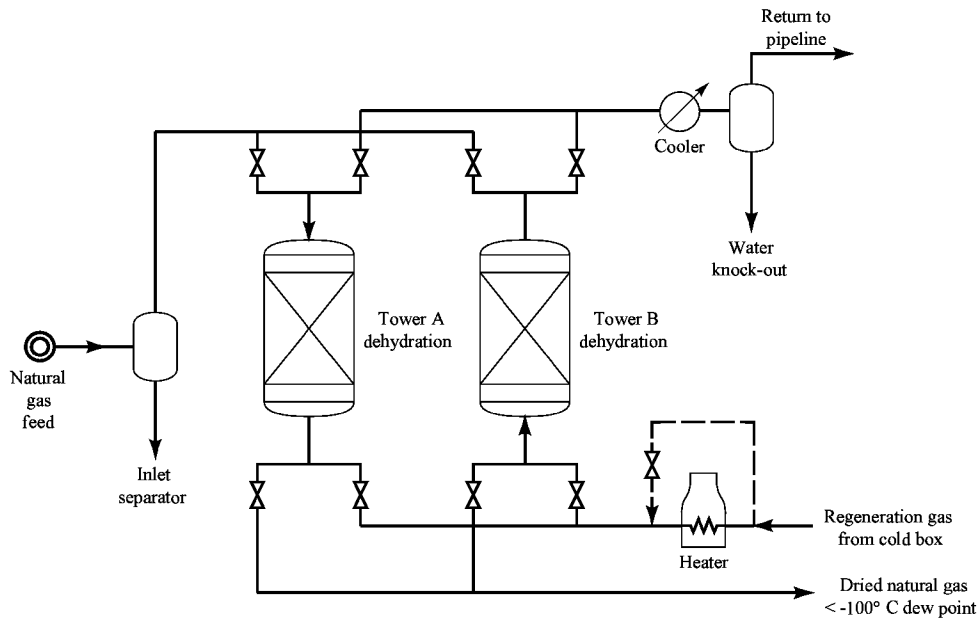


Fig. 12. Molecular sieve natural gas dehydrator.

Single-bed systems can also be used if the demand for drying is intermittent. For example, pressure-swing air dryers installed on heavy trucks usually employ a single desiccant bed to dry compressed air for operating the brakes. When the pressure in the primary air storage reservoir in the system reaches a set point, a blowdown to atmospheric pressure and a purge of the dryer bed using air from a purge reservoir take place (32). A single-bed, thermal-swing system can also be used for drying intermittent flows. Very dry gas is often needed to regenerate or cool adsorbent beds for gas purification (see ADSORPTION, GAS SEPARATION). A single-bed auxiliary dryer can be used for this purpose (33).

Three-bed solid adsorption systems are also used for drying gases. The three beds can be arranged so that two beds are drying and one is in regeneration or one bed is drying and two are in regeneration, one in heating and one in cooling. When two beds are used in drying, they can be in series flow, one bed in the "lead" position and the other in the "lag" or "trim" position, or they can be in parallel flow, where the adsorption fronts are staggered.

In designing a gas drying system, the engineer must often estimate the water content of the gas to be dried. If the gas is air at atmospheric pressure, psychrometric charts are available to give the specific humidity (mass of water per mass of dry air) for given conditions of dry bulb temperature and relative humidity or wet bulb temperature (34,35). Often, however, the gas is under pressure and is known to be saturated with water under given temperature and pressure conditions. In this case, an estimate of the water content can be made by using the vapor pressure of water p° at the given temperature at total pressure π , as reported in the steam tables.

$$y = \frac{p^\circ}{\pi} \tag{8}$$

Although this estimate will sometimes be sufficient, it may well fail at higher pressures (36). Fortunately, several references are available that provide data and estimating procedures for high pressure hydrocarbons (C_1 – C_4), natural gas, hydrogen, and nitrogen (37–54).

Adsorption Plots. Isotherm plots are the most common method of presenting adsorption data. An isotherm is a curve of constant temperature; the adsorbed water content of the adsorbent is plotted against the water partial pressure in equilibrium with the adsorbent. An isostere plot shows curves of constant adsorbed water content: the vapor pressure in equilibrium with the adsorbent is plotted against temperature. Figure 13 shows isosteres for the three primary adsorbents described previously. In this case, the dew points for the three adsorbents are plotted at 0.5, 5, and 10 kg

H₂O 100 kg of adsorbent. At equilibrium and at a given adsorbed water content, the dew point that can be obtained in the treated fluid is a function only of the adsorbent temperature. The slopes of the isosteres indicate that the capacity of molecular sieves is less temperature sensitive than that of silica gel or activated alumina. In another type of isostere plot, the natural logarithm of the vapor pressure of water in equilibrium with the desiccant is plotted against the reciprocal of absolute temperature. The slopes of these isosteres are proportional to the isosteric heats of adsorption of water on the desiccant (see ADSORPTION, GAS SEPARATION).

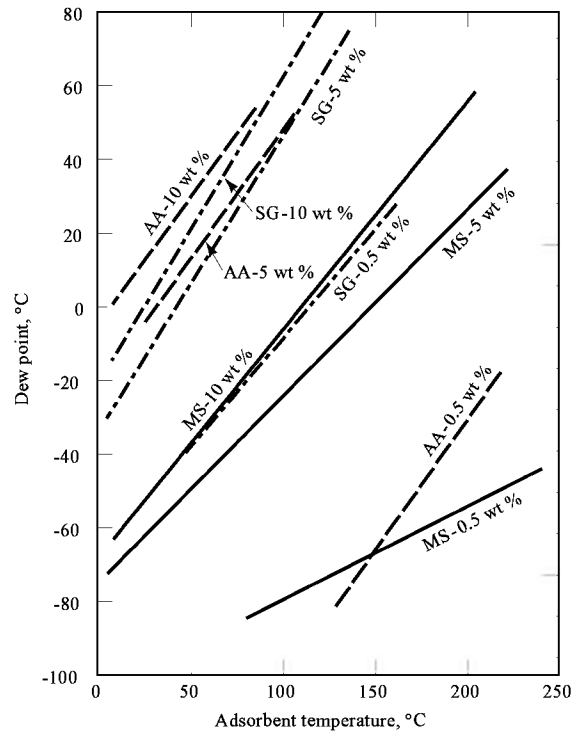


Fig. 13. Adsorbent isosteres for activated alumina (AA), silica gel (SG), and molecular sieves (MS).

Mass Transfer and Useful Capacity. The term useful capacity, also referred to earlier as breakthrough capacity, differs from the equilibrium capacity shown on Figures 9a and 9b. The useful capacity is a measure of the total moisture taken up by a packed bed of adsorbent at the point where moisture begins to appear in the effluent. Thus the drying process cycle must be stopped before the adsorbent is fully saturated. The portion of the bed that is not saturated to an equilibrium level is called the mass-transfer zone.

The parameters affecting the size and shape of a mass-transfer zone are adsorbent type, adsorption isotherm shape, flow rate, packed-bed depth, adsorbent particle size, physical properties of the carrier fluid, temperature, pressure, and the concentration of water in the carrier fluid. For example, as the particle size increases for a given fluid flow rate, the pressure drop decreases. However, as the particle size increases, the mass-transfer resistance increases, and the adsorbent takes longer to achieve its equilibrium water capacity. As a result, larger amounts of desiccant are required. Therefore, the optimal particle size is a compromise between pressure drop (energy consumption) and desiccant utilization. If conditions are chosen that are favorable to mass transfer, eg, small particle size, then the mass-transfer zone is small when compared to the total amount of packed bed employed in drying service. In this case, bed utilization is high and the breakthrough or useful capacity closely approaches the true equilibrium capacity. More often, conditions cannot be optimized on the basis of adsorbent needs but are fixed by the needs of the drying process. This situation may dictate unfavorable mass-transfer conditions when pressure drop is limited or practical packed-bed diameters and depths must be employed. Bed utilization is then reduced and the breakthrough capacity falls short of the equilibrium capacity.

Figure 14 depicts the location of the water front in packed beds of adsorbents at a short but typical contact time for dehydration. The percentage of useful capacity as compared to equilibrium capacity is about 50% for activated alumina and silica gel and more than 90% for the molecular sieve. The reason for this difference in performance lies in the shape of the isotherm. Under the Brunauer classification, the molecular sieve has a type I isotherm shape, which is favorable for adsorption mass transfer (see ADSORPTION). This isotherm produces a self-sharpening adsorption front, or constant pattern behavior. Silica gel and activated alumina have non-type I isotherms, which are unfavorable for adsorption mass transfer. Their isotherms produce an expanding adsorption front or proportional pattern behavior. Thus the highest equilibrium capacity adsorbent may not always result in the highest useful capacity in a dynamic system. If higher moisture contents can be tolerated in the effluent, then a larger fraction of the mass-transfer zone can be allowed to leak through into the effluent stream. This technique improves bed utilization at the expense of an increase in the effluent dew point.

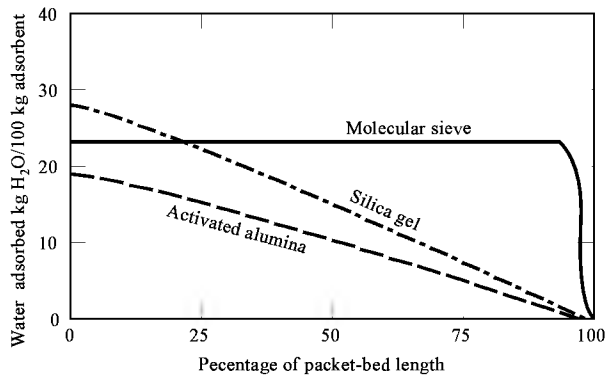


Fig. 14. Position of water front in packed bed of adsorbent during dynamic dehydration. Conditions: 50% rh; 10.2 cm³ s air; particle size ca0.167 cm; temperature 25°C; contact time 1.7 s.

Economic Aspects

Cost-Effectiveness of Desiccants. The cost-effectiveness of a regenerable Class 4 (physical adsorption) desiccant is a function of the initial investment and the operating cost. The initial investment depends on the desiccant bed weight, which determines the costs of the desiccant and the desiccant vessel. The investment also includes other equipment costs, such as compressors or blowers for gas-phase systems, and regeneration equipment, such as regeneration compressor and heater. The principal operating costs are the energy costs associated with thermal-swing regeneration and compression costs associated with fluid pressure drop across the bed.

The required desiccant weight is a function of several factors: the water removal requirements (mass time), the cycle time, the equilibrium loading of water on the desiccant at the feed conditions, the residual water loading on the desiccant after regeneration, and the size of the mass-transfer zone of the desiccant bed. These factors, in turn, depend on the flow rate, temperature, pressure, and water content of both the fluid being dried and the regeneration fluid (see ADSORPTION, GAS SEPARATION).

Shorter cycle times produce smaller bed sizes. The minimum cycle time is usually dictated by the minimum regeneration time required to heat and cool the bed. Systems of greater than two beds provide some flexibility in regeneration time but add to investment costs.

Operating costs consist of compression costs for drying gases and the energy costs for thermally regenerated systems. Labor is also a cost, although most systems are designed to run automatically with a minimum of operator attention. Compression costs are a function of bed size and design and desiccant particle characteristics as well as flow rate, pressure, and temperature. Regeneration costs in thermal-swing systems include the sensible heating of the desiccant and vessel from the feed temperature to the regeneration temperature and the heat of adsorption of water on the desiccant. Some designs with two beds in regeneration (one in heating and one in cooling) allow for recovery of sensible heat.

Desiccant replacement is another operating cost, although desiccant life is usually several years in regenerable applications. Desiccant life depends on the stability of the desiccant in the given service, the frequency of regeneration, the presence of reactive contaminants, and the possibility of upsets to normal operation.

Market Data. The largest U.S. manufacturer of molecular sieves for adsorbent and desiccant use is UOP, which has a production capacity of 18–20 million kg year. W.R. Grace and Zeochem have about 7 and 2 million kg year capacity, respectively (55). W.R. Grace is the largest producer of silica gel desiccants. Activated alumina for use as adsorbent and desiccant is produced by LaRoche Chemicals (formerly Kaiser) and by Aluminum Company of America. About one-third of the U.S. supply of activated alumina adsorbent and desiccant is imported by Rhøfne-Poulenc.

The largest users of molecular sieve desiccants are the natural gas processing, insulating glass, and refrigeration industries (Table 4). Much smaller quantities of silica gel are used in these applications. Silica gel dominates the packaging industry, where the material is used to protect electronic equipment and pharmaceuticals, for example, from moisture. Silica gel is also used in dehumidification of buildings. The principal uses of activated alumina are in refrigeration, where its primary function is to adsorb organic acids rather than water, air drying, and alkylaton feed stream drying in oil refineries. The data in Table 4 are estimates of 1990 U.S. usage and are considered the best available (56).

Table 4. Applications of Primary Class 4 Desiccants^a

Application	Desiccants, t		
	Molecular sieve	Silica gel	Activated alumina
natural gas processing	5500	1200	230
insulating glass	3900	1300	
refrigeration	3900	320	2300
packaging	small	1400	
air and inert gas drying	180	410	3400
alkylation	small		2700

^a Estimated 1990 U.S. usage.

Future Applications

Energy Storage. Reactivating a desiccant stores the reactivation energy in the dehydrated desiccant. This energy-storage feature is useful if the energy source is intermittent or seasonal, such as solar energy, or interruptable. Research suggests that this energy-storage feature is especially useful if the desiccant is used to dry air for agricultural applications (57).

Desiccant Cooling.

Considerable work is now being done in the field of desiccant-based cooling systems. In these systems, a desiccant is used to produce an extremely dry air stream. The dry air is then cooled in a heat exchanger and humidified with a water spray. The evaporation of the water absorbs heat from the air and produces a cold, almost saturated air stream for air conditioning. The desiccant is thermally regenerated with exhaust air, which is heated with solar energy, natural gas, or waste heat from power generation. Both liquid (58) and solid desiccant systems have been studied. The solid desiccant systems typically employ a wheel, which rotates a bed of desiccant particles through sectors in which drying, thermal regeneration, and cooling take place (59). These systems are extensions of existing dehumidification technology, with a strong emphasis placed on minimizing the thermal regeneration requirements. To this end, extensive basic studies have been carried out of solid desiccants and the properties of the hypothetical desiccant ideally suited to low temperature thermal regeneration. Research shows that the shape of the desiccant isotherm and the thermal properties of the desiccant, such as heat capacity, are equally important in desiccant selection (60). Organic polymers as well as inorganic materials are being considered as desiccant materials (61).

BIBLIOGRAPHY

"Drying Agents" in *ECT* 1st ed., Vol. 5, pp. 266–276, by J. F. Skelly, The M. W. Kellogg Co.; in *ECT* 2nd ed., Vol. 7, pp. 378–398, by B. K. Beecher, Wyandotte Chemicals Corp.; in *ECT* 3rd ed., Vol. 8, pp. 114–130, by J. P. Ausikaitis, Union Carbide Corp.

1. J. J. Collins, *Chem. Eng. Prog-Sym. Ser.* **63**(74), 31 (1967).
2. A. D. F. Toy and E. N. Walsh, *Phosphorus Chemistry in Everyday Living*, 2nd ed., American Chemical Society, Washington, D.C., 1987, pp. 233–235.
3. A. P. Cohen, "Test Methods for the Compatibility of Desiccants with Alternative Refrigerants," *ASHRAE Transactions*, Vol. 99, Part 1, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, Ga., 1993.
4. O. A. Hougen and F. W. Dodge, *The Drying of Gases*, J. W. Edwards, Ann Arbor, Mich., 1947, p. 30.
5. R. S. Boynton, *Chemistry and Technology of Lime and Limestone*, John Wiley & Sons, Inc., New York, 1966.
6. *Drierite Product Catalog*, W. A. Hammond Drierite Co., Xenia, Ohio.
7. M. Shepherd, "Relative Efficiencies of Drying Agents," *International Critical Tables*, Vol. 3, McGraw-Hill, Inc., New York, 1928.
8. K. G. Davis and K. D. Manchanda, *Chem. Eng.* **81**, 102 (Sept. 16, 1974).
9. D. R. Burfield, K. H. Lee, and R. H. Smithers, *J. Org. Chem.* **42**(18), 3060–3065 (1977). Part 1 of a series of nine papers. Contains data on efficiency of drying benzene, 1,4-dioxane, and acetonitrile with 4A molecular sieve, Al₂O₃, silica gel, CaH₂, LiAlH₄, Na, Na–K alloy, P₂O₅, CaCl₂, Na₂SO₄, KOH, CaSO₄, and MgSO₄.

10. D. R. Burfield, G. H. Gan, and R. H. Smithers, *J. Appl. Chem. Biotechnol.* **28**, 23 (1978). Part 2, efficiency of drying 1,4-dioxane, tetrahydrofuran, acetonitrile, and benzene with 3A and 4A molecular sieves.
11. D. R. Burfield and R. H. Smithers, *J. Org. Chem.* **43**(20), 3966–3968 (1978). Part 3, efficiency of drying acetone, dimethylformamide, dimethylsulfoxide, and hexamethylphosphoric triamide with Al_2O_3 , BaO, B_2O_3 , CaH_2 , CaO, CaSO_4 , CuSO_4 , KOH, K_2CO_3 , molecular sieves 3A and 4A, Na–K alloy, and P_2O_5 .
12. D. R. Burfield and R. H. Smithers, *J. Chem. Tech. Biotechnol.* **30**, 491–496 (1980). Part 4, efficiency of drying 1,4-dioxane, toluene, and dichloromethane with cationic exchange resins.
13. D. R. Burfield, R. H. Smithers, and A. S. C. Tan, *J. Org. Chem.* **46**, 629–631 (1981). Part 5, efficiency of drying pyridine, 2-methylpyridine, 2,6-dimethylpyridine, 2,4,6-trimethylpyridine, triethylamine, diisopropylamine, and 1,3-propanediamine with Al_2O_3 , BaO, CaC_2 , CaH_2 , CaO, CaSO_4 , KOH, molecular sieves 3A and 4A, Na, and silica gel.
14. D. R. Burfield and R. H. Smithers, *J. Chem. Educ.* **59**(8), 703–704 (1982). Part 6, efficiency of drying water-saturated diethylether with MgSO_4 , CaSO_4 , CaCl_2 , cation-exchange resin, molecular sieves 4A and 5A, Na_2SO_4 , K_2CO_3 , polyacrylamide, and Al_2O_3 .
15. D. R. Burfield, R. H. Smithers, and A. S. C. Tan, *J. Org. Chem.* **48**, 2420–2422 (1983). Part 7, efficiency of drying methanol, ethanol, 2-butanol, *t*-butyl alcohol, and 1,2-ethanediol with molecular sieves 3A and 4A, MgSO_4 , CaC_2 , B_2O_3 , BaO, CaO, Mg, Al, and benzene azeotrope.
16. D. R. Burfield, G. T. Hefter, and D. S. P. Koh, *J. Chem. Tech. Biotechnol.* **34A**, 187–194 (1984). Part 8, drying with molecular sieve 3A.
17. D. R. Burfield, *J. Org. Chem.* **49**, 3852–3854 (1984). Part 9, efficiency of drying 1,4-dioxane, water-saturated diethylether, and acetonitrile with CaSO_4 , CaCl_2 , and molecular sieves 3A and 4A.
18. A. P. Cohen and S. R. Dunne, "Review of Automotive Air-Conditioning Drydown Rate Studies: The Kinetics of Drying Refrigerant 12," *ASHRAE Transactions*, Vol. 93, Part 2, American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Atlanta, Ga., 1987.
19. S. R. Dunne, *Automotive Corrosion and Prevention Conference Proceedings*, Society of Automotive Engineers, Warrendale, Pa., 1989, pp. 165–173.
20. I. Mellan, *Polyhydric Alcohols*, Spartan Books, Washington, D.C., 1962.
21. G. O. Curme, Jr. and F. Johnston, eds., *Glycols*, American Chemical Society Monograph No. 114, Rheinhold Publishing Corp., New York, 1953.
22. L. D. Polderman, "The Glycols as Hydrate Point Depressants in Natural Gas Systems," *Proceedings of the Gas Conditioning Conference*, University of Oklahoma, Norman, 1958.
23. "Triethylene Glycol," Technical Bulletin F-49191A-ICD, Union Carbide Corp., Danbury, Conn., 1989.
24. "Gas Treating Chemicals," Technical Bulletin F-4 1335C, Union Carbide Chemicals and Plastics Co., Danbury, Conn., 1980.
25. "A Guide to Glycols," Technical Bulletin 117-00991-4 91, The Dow Chemical Company, Midland, Mich., 1981.
26. C. Misra, *Industrial Alumina Chemicals*, ACS Monograph 184, American Chemical Society, Washington, D.C., 1986, pp. 107–120.
27. R. K. Iler, *The Chemistry of Silica*, John Wiley & Sons, Inc., New York, 1979, pp. 462–621.
28. D. W. Breck, *Zeolite Molecular Sieves: Structure, Chemistry, and Use*, Wiley-Interscience, New York, 1974, 715 pp.
29. P. N. Kraychy and A. Masuda, *Oil Gas J.* **64** (1966).
30. D. Basmadjian, *Adv. Drying*, **3**, 307–357 (1984). A review of thermal-swing industrial drying applications with 108 references.
31. A. L. Weiner, *Chem. Eng.* **81**, 92 (1974).
32. J. P. Ausikaitis, "Pneumatic System Air Drying by Pressure Swing Adsorption," *ACS Symposium Series, No. 40, Molecular Sieves-II*, American Chemical Society, Washington, D.C., 1977.
33. **U.S. Pat. 4,484,933** (Nov. 27, 1984), A. P. Cohen (to Union Carbide Corp.).
34. *ASHRAE Handbook 1989 Fundamentals*, American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Atlanta, Ga., 1989, Chapt. 6.
35. *Perry's Chemical Engineers' Handbook*, 6th ed., McGraw-Hill, New York, 1989, p. 12-4.
36. S. B. Adler, H. Ozkardesh, and C. F. Spencer, *Oil Gas J.*, 107–113 (Nov. 17, 1980).
37. J. M. Campbell, *Gas Conditioning and Processing*, 3rd ed., Campbell Petroleum Series, Norman, Okla., 1974, 266 pp.
38. S. C. Sharma, "Equilibrium Water Content of Gaseous Mixtures," Ph.D. dissertation, University of Oklahoma, Norman, 1969. Available from University Microfilms, Ann Arbor, Mich., order #69-8601.
39. R. G. Anthony and J. J. McKetta, *J. Chem. Eng. Data* **12**(1), 17 (1967).
40. Ref. 39, p. 21.
41. R. G. Anthony and J. J. McKetta, *Hydrocarbon Process.* **47**(6), 131 (1968).
42. E. P. Bartlett, *J. Am. Chem. Soc.* **49**, 65 (1927).
43. R. Kobayashi and D. L. Katz, *Ind. Eng. Chem.* **45**(2), 440 (1953).
44. C. C. Li and J. J. McKetta, *J. Chem. Eng. Data* **8**(2), 247 (1963).
45. J. J. McKetta and D. L. Katz, *Ind. Eng. Chem.* **40**(5), 853 (1948).
46. R. H. Olds, B. H. Sage, and W. N. Lacey, *Ind. Eng. Chem.* **34**(10), 1223 (1942).
47. F. Pollitzer and E. Strebel, *Z. Phys. Chem.* **110**, 768 (1924).
48. H. H. Reamer, R. H. Olds, B. H. Sage, and W. N. Lacey, *Ind. Eng. Chem.* **35**(7), 790 (1943).
49. H. H. Reamer, R. H. Olds, B. H. Sage, and W. N. Lacey, *Ind. Eng. Chem.* **36**(4), 381 (1944).
50. M. Rigby and J. M. Prausnitz, *J. Phys. Chem.* **72**(1), 330 (1968).
51. J. N. Robinson, "Estimation of the Water Content of Sour Natural Gas," presented at *Canadian Natural Gas Processing Association*, Calgary, Alberta, Mar. 9, 1978.
52. A. W. Saddington and N. W. Krase, *J. Am. Chem. Soc.* **56**, 353 (1934).
53. A. H. Wehe and J. J. McKetta, *J. Chem. Eng. Data* **6**(2), 167 (1961).
54. J. J. McKetta and A. H. Wehe, *Petrol. Ref.* **37**(8), (1958). Data and estimating procedures for saturated water content of natural gas.
55. M. Smart, "Zeolites," *Chemical Economics Handbook*, SRI International, Menlo Park, Calif., 599.1000, 1992.
56. *Adsorbents & Desiccants, Markets and Technology, 1983–1990*, Oxenham Technology Associates, Inc., New York, 1985.
57. W. M. Miller, *Ener. Ag.* **2**, 341–354 (1983).
58. J. R. Howell, *NATO ASI Ser. Ser E* **129**, 374–387 (1987).
59. L. G. Harriman, III, ed., *The Dehumidification Handbook*, 2nd ed., Munters Cargocaire, Amesbury, Mass., 1990, pp. 3–14 to 3–18.
60. R. K. Collier, Jr., *Advanced Desiccant Materials Assessment, Phase II, Final Report*, Gas Research Institute, GRI 88 0125, Chicago, Jan. 1988.
61. A. W. Czandema, *Polymers as Advanced Materials for Desiccant Applications: 1987*, Solar Energy Research Institute, Golden, Colo., 1988.

Alan P. Cohen
UOP

DESIGN OF EXPERIMENTS

Obtaining valid results from a test program calls for commitment to sound statistical design. In fact, proper experimental design is often more important than sophisticated statistical analysis. Results of a well-planned experiment are often evident from simple graphical analyses. However, the world's best statistical analysis cannot rescue a poorly planned experimental program. The main reason for designing an experiment statistically is to obtain unambiguous results at a minimum cost. The need to learn about interactions among variables and to measure experimental error are added reasons.

Many chemists and engineers think of experimental design mainly in terms of standard plans for assigning treatments to experimental units, such as the Latin square, factorial, fractional factorial, and central composite (or Box) designs. These designs are described in books, such as those summarized in the General References, and catalogued in various reports and articles. Important as such formal plans are, the final selection of test points represents only the proverbial tip of the iceberg, the culmination of a careful planning process.

Statistically planned experiments are characterized by (1) the proper consideration of extraneous variables; (2) the fact that primary variables are changed together, rather than one at a time, in order to obtain information about the magnitude and nature of interactions and to gain improved precision in estimating the effects of these variables; and (3) built-in procedures for measuring the various sources of random variation and for obtaining a valid measure of experimental error against which one can assess the impact of the primary variables and their interactions.

A well-planned experiment is often tailor-made to meet specific objectives and to satisfy practical constraints. The final plan may or may not involve a standard textbook design. If possible, a statistician knowledgeable in the design of experiments should be called in early, made a full-fledged team member, and be fully apprised of the objectives of the test program and of the practical considerations and constraints. He or she may contribute significantly merely by asking probing questions. After the problem and the constraints have been clearly defined, the statistician can evolve an experimental layout to minimize the required testing effort to obtain the desired information. This may or may not involve a standard statistical design. However, designing an experiment is often an iterative process requiring rework as new information and preliminary data become available. With a full understanding of the problem, the statistician is in an improved position to respond rapidly if last-minute changes are required and to provide meaningful analyses of the subsequent experimental results.

Much of this article previously appeared in References 1 and 2. Reference 3 proposes a systematic approach to planning a designed experiment. References 4–7 illustrate many of the considerations discussed here and introduce a few others. References 8 and 9 provide two other surveys dealing with the design of experiments, including more detailed discussions than given here of specific designs and their analysis. Reference 10 provides a relatively technical overview of experimental design concepts and plans. In addition, the books listed and described in the General References provide information concerning the technical aspects of experimental design.

Purpose and Scope of the Experiment

Designing an experiment is like designing a product. Every product serves a purpose; so should every experiment. This purpose must be clearly defined at the outset. It may, for example, be to optimize a process, to estimate the probability that a component operates properly under a given stress for a specified number of years, to evaluate the relative effects on product performance of different sources of manufacturing and end use variability, or to determine whether a new process is superior to an existing one. An understanding of this purpose is important in developing an appropriate experimental plan.

In addition to defining the purpose of a program, one must decide on its scope. An experiment is generally a vehicle for drawing inferences about the real world. Since it is highly risky to draw inferences about situations beyond the scope of the experiment, care must be exercised to make this scope sufficiently broad. This must especially be kept in mind if laboratory findings are to apply to a (possibly future) manufacturing line. Thus if the results are material dependent, the material for fabricating experimental units must be representative of what one might expect to encounter in production. If the test program were limited to a single heat of steel or a single batch of raw material, the conclusions might be applicable only to that heat or batch, irrespective of the sample size. Similarly, in deciding whether or not temperature should be included as an experimental variable to compare different preparations, it must be decided whether the possible result that one preparation outperforms another at a particular temperature would also apply for other temperatures of practical interest. If not, one should seriously consider including temperature as an experimental variable.

Experimental Variables

An important part of planning an experimental program is the identification of the variables that affect the response and deciding what to do about them. The decision as to how to deal with each of the candidate variables can be made jointly by the experimenter and the statistician. However, identifying the variables is the experimenter's responsibility. Controllable or independent variables in a statistical experiment can be dealt with in four different ways. The assignment of a particular variable to a category often involves a trade-off among information, cost, and time.

Primary Variables. The most obvious variables are those whose effects on performance are to be evaluated directly; these are the variables that, most likely, created the need for the investigation in the first place. Such variables may be quantitative, such as catalyst concentration, temperature, or pressure, or they may be qualitative, such as method of preparation, catalyst type, or batch of material.

Quantitative controllable variables are frequently related to the response (or performance) variable by some assumed statistical relationship or model. The minimum number of conditions or levels per variable is determined by the form of the assumed model. For example, if a straight-line relationship can be assumed, two levels (or conditions) may be sufficient; for a quadratic relationship a minimum of three levels is required. However, it is often desirable to include some added points, above the minimum needed, so as to allow assessment of the adequacy of the assumed model.

Qualitative variables can be broken down into two categories. The first consists of those variables whose specific effects are to be compared directly; eg, comparison of the effect on performance of two proposed preparation methods or of three catalyst types. The required number of conditions for such variables is generally evident. Such variables are sometimes referred to as fixed effects or Type I variables.

The second type of qualitative variables are those whose individual contributions to performance variability are to be evaluated. The specific conditions of such variables are randomly determined. Material batch would be such a variable if one is not interested in the behavior of specific batches per se, but instead needs information concerning the magnitude of variation in performance caused by differences among batches. In this case, batches would be selected randomly from a large population of batches. For variables of this type, it is generally desirable to have a reasonably large sample (eg, five or more) so as to obtain an adequate degree of precision in estimating the variability in response attributable to the variable. Such variables are sometimes referred to as random effects or Type II variables.

When there are two or more variables, they might interact with one another, ie, the effect of one variable upon the response depends on the value of the other variable. Figure 1 shows a situation where two noninteracting variables, preparation type and temperature, independently affect time to rupture, ie, the effect of temperature on time to rupture is the same for both preparation types. In contrast, Figure 2 shows two examples of interactions between preparation and temperature.

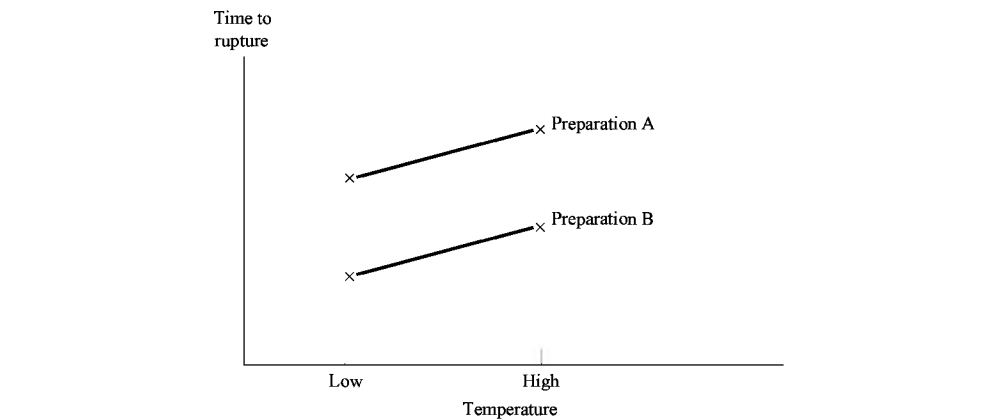


Fig. 1. Situation with no interaction between variables.

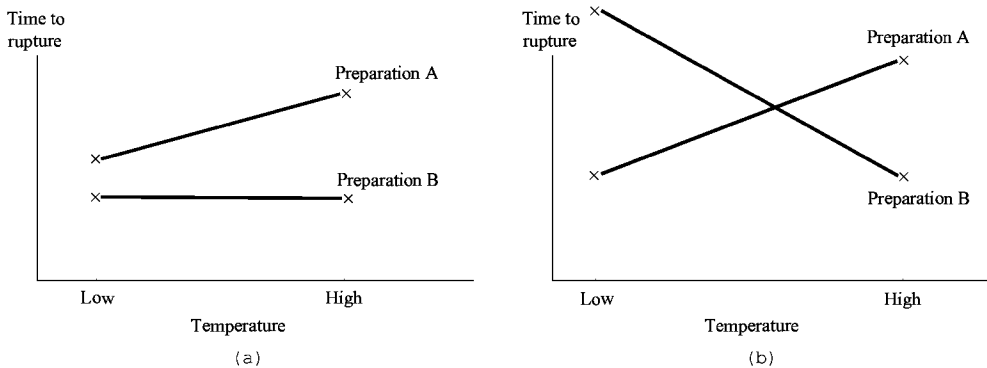


Fig. 2. Situation with interactions between variables, where in (a) an increase in temperature is beneficial for preparation A but does not make any difference for preparation B, and (b) an increase in temperature raises time to rupture for preparation A but decreases it for preparation B.

An important purpose of a designed experiment is to obtain information about interactions among the primary variables. This is accomplished by varying factors simultaneously rather than one at a time. Thus in Figure 2, each of the two preparations would be run at both low and high temperatures using, for example, a full factorial experiment.

Background Variables and Blocking. In addition to the primary controllable variables there are those variables, though not of primary interest, that cannot, and perhaps should not, be held constant. Such variables arise, for example, when an experiment is to be run over a number of days, or when different machines or operators or both are to be used. It is of crucial importance that such background variables are not varied in complete conjunction with the primary variables. For example, if preparation A were run only on day 1 and preparation B only on day 2, it is impossible to determine how much of any observed difference in response between the two preparations is the result of normal day-to-day process variation. Such mixing of effects is called confounding. If the background variables do not interact with the primary variables, they may be introduced into the experiment in the form of experimental blocks. An experimental block represents a relatively homogeneous set of conditions within which different conditions of the primary variables are compared.

A well-known example of blocking arises in the comparison of wear for different types of automobile tires. Tire wear may vary from one automobile to the next, irrespective of the tire type, because of differences among automobiles, variability among drivers, and so forth. Assume, for example, that for the comparison of four tire types (A, B, C, and D), four automobiles (1, 2, 3, and 4) are available. A poor procedure would be to use the same type of tire on each of the four wheels of an automobile and vary the tire type among automobiles, as in the following tabulation:

Automobile			
1	2	3	4
A	B	C	D
A	B	C	D
A	B	C	D
A	B	C	D

Such an assignment is undesirable because the differences between tires cannot be separated from the differences between automobiles in the subsequent analysis. Separation of these effects can be obtained by treating automobiles as experimental blocks and randomly assigning tires of each of the four types to each automobile as follows:

Automobile			
1	2	3	4
A	A	A	A
B	B	B	B
C	C	C	C
D	D	D	D

The above arrangement is known as a randomized block design.

The symmetry of the preceding example is not always found in practice. For example, there may be six tire types under comparison and fifteen available automobiles. Tires are then assigned to automobiles to obtain the most precise comparison among tire types, using a so-called incomplete block design. Similar concepts apply if there are two or more primary variables, rather than tire type alone.

A main reason for running an experiment in blocks is to ensure that the effect of a background variable does not contaminate evaluation of the effects of the primary variables. However, blocking removes the effect of the blocked variables from the experimental error as well, thus allowing more

precise estimation of the experimental error and, as a result, more precise estimates of the effects of the primary variables. Finally, in many situations, the effect of the blocking variables on the response can also be readily evaluated, a further advantage of blocking.

In some situations, there may be more than one background variable whose possible contaminating effect is removed by blocking. Thus in the automobile tire comparison, the differences between wheel positions may be of concern in addition to differences between automobiles. In this case, wheel position might be introduced into the experiment as a second blocking variable. If there are four tire types to be compared, this might, for example be done by randomly assigning the tires of each of the four types according to the following plan, known as a Latin square design:

Wheel position	Automobile			
	1	2	3	4
1	A	D	C	B
2	B	A	D	C
3	C	B	A	D
4	D	C	B	A

In this plan, the effects of both automobile and wheel position are controlled by blocking. It should, however, be kept in mind that for the Latin square design, as for other blocking plans, it is generally assumed that the blocking variables do not interact with the primary variable to be evaluated.

Uncontrolled Variables and Randomization. A number of further variables, such as ambient conditions (temperature, pressure, etc), can be identified but not controlled, or are only hazily identified or not identified at all but affect the results of the experiment. To ensure that such uncontrolled variables do not bias the results, randomization is introduced in various ways into the experiment to the extent that this is practical.

Randomization means that the sequence of preparing experimental units, assigning treatments, running tests, taking measurements, and so forth, is randomly determined, based, for example, on numbers selected from a random number table. The total effect of the uncontrolled variables is thus lumped together into experimental error as unaccounted variability. The more influential the effect of such uncontrolled variables, the larger the resulting experimental error, and the more imprecise the evaluations of the effects of the primary variables. Sometimes, when the uncontrolled variables can be measured, their effect can be removed from experimental error statistically.

Background variables could also be introduced into the experiment by randomization, rather than by blocking techniques. Thus in the previous example, the four tires of each type could have been assigned to automobiles and wheel positions completely randomly, instead of treating automobiles and wheel positions as experimental blocks. This could have resulted in an assignment such as the following:

Wheel position	Automobile			
	1	2	3	4
1	B	D	B	D
2	A	D	D	A
3	C	C	A	D
4	B	B	C	A

Both blocking and randomization generally ensure that the background variables do not contaminate the evaluation of the primary variables. Randomization sometimes offers the advantage of greater simplicity compared to blocking. However, under blocking the effect of a background variable is removed from the experimental error, whereas under randomization it usually is not. Thus the aim might be to remove the effects of the one or two most important background variables by blocking, while counteracting the possible contaminating effect of others by randomization.

Variables Held Constant. Finally, some variables should be held constant in the experiment. Holding a variable constant limits the size and complexity of the experiment but, as previously noted, can also limit the scope of the resulting inferences. The variables to be held constant in the experiment must be identified and the mechanisms for keeping them constant defined. The experimental technique should be clearly specified at the outset of the experiment and closely followed.

Experimental Environment and Constraints

The operational conditions under which the experiment is to be conducted and the manner in which each of the factors is varied must be clearly spelled out.

Variables are not created equal; some can be varied more easily than others. For example, a change in pressure may be implemented by a simple dial adjustment; on the other hand, stabilization requirements might make it difficult to change temperature. In such situations, completely randomizing the sequence of testing is impractical. However, basing the experimental plan on convenience alone may lead to ambiguous and unanalyzable results. For example, if the first half of the experiment is run at one temperature and the second half at another temperature, it may not be possible to tell whether the observed difference in response results from the difference in temperature or from some other factors that varied during the course of the experiment such as raw material, ambient conditions, or operator technique. Thus the final experimental plan must be a compromise between cost and information. The experiment must be practical to run, yet still yield statistically valid results (11,12).

Practical considerations enter into the experimental plan in various other ways. In many programs, variables are introduced at different operational levels. For example, in evaluating the effect of alloy composition, oven temperature, and varnish coat on tensile strength, it may be convenient to make a number of master alloys with each composition, split the alloys into separate parts to be subjected to different heat treatments, and then cut the treated samples into subsamples to which different coatings are applied. Tensile strength measurements are then obtained on all coated subsamples.

Situations such as the preceding arise frequently in practice and are referred to as split-plot experiments. The terminology results from the agricultural origins of experimental design, eg, a farmer needed to compare different fertilizer types on a plot of land with varying normal fertility. A characteristic of split-plot plans is that more precise information is obtained on the low level variables (varnish coats in the preceding example) than on the high level variables (alloy composition). The split-plot nature of the experimental environment, if present, is important information, both in the planning and in the analysis of the experiment.

Prior Knowledge. Prior knowledge is often available concerning the expected outcome at certain experimental conditions. For example, some combinations of conditions might be known to yield poor results or might not be attainable with the equipment available, or worse yet, could result in blowing up the plant. Furthermore, all proposed conditions in the experiment need to make sense. For example, a misguided proposed experiment had a condition that required a resistance of 50 ohms for a circuit without resistors. Clearly unreasonable or hazardous conditions must be omitted from the experimental design, irrespective of whether they happen to coincide with a standard statistical pattern. Thus the experiment must be adjusted to accommodate reality and not the reverse.

Response Variables. A clear statement is required of the performance characteristics or dependent variables to be evaluated as the experimental response. Even a well-designed experiment fails if the response cannot be measured properly. Frequently, there are a number of response variables; for example, tensile strength, yield strength, % elongation, etc. It is important that standard procedures for measuring each variable be established and documented. Sometimes the response is on a semiquantitative scale; eg, material appearance may be in one of five categories, such as outstanding, superior, good, fair, and poor. In this case, it is particularly important that standards that define (and perhaps illustrate) each of these categories are developed initially, especially if judgments are to be made at different times and perhaps by different observers.

Types of Repeat Information. The various ways of obtaining repeat results in the experiment need to be specified. Different information

about repeatability is obtained by (1) taking replicate measurements on the same experimental unit, (2) cutting a sample in half at the end of the experiment and obtaining a reading on each half, and (3) taking readings on two samples prepared independently of one another, eg, on different runs, at the same aimed-at conditions. Frequently, there is greater homogeneity among replicate measurements on the same sample than among measurements on different samples. The latter reflect the random unexplained variation of repeat runs conducted under identical conditions. A skillfully planned experiment imparts information about each component of variability, if such information is not initially available, and uses a mixture of replication and repeat runs to yield the most precise information for the available testing budget. The way in which such information is obtained must also be known to perform a valid analysis of the results.

Preliminary Estimates of Repeatability. Initial estimates of repeatability should be obtained before embarking on any significant test program. Such information may be available from previous testing; if it is not, valid preliminary runs should be conducted at different times under supposedly identical conditions. If these runs result in large variability in performance, the important variables that affect the results have not been identified, and further research may be needed before the proposed experiment can commence.

Consistent Data-Recording Procedures. Clear procedures for recording all pertinent data from the experiment must be developed and documented, and unambiguous data recording forms established. These should include provisions not only for recording the values of the measured responses and the desired experimental conditions, but also the conditions that resulted, if these differ from those planned. It is generally preferable to use the values of the actual conditions in the statistical analysis of the experimental results. For example, if a test was supposed to have been conducted at 150°C but was run at 148.3°C, the actual temperature would be used in the analysis. In experimentation with industrial processes, process equilibrium should be reached before the responses are measured. This is particularly important when complex chemical reactions are involved.

The values of any other variables that might affect the responses should also be recorded, if possible. For example, although it may not be possible to control the ambient humidity, its value should be measured if it might affect the results. In addition, variations in the factors to be held constant, special happenings (eg, voltage surges), and other unplanned events should be recorded. The values of such covariates can be factored into the statistical analysis, thereby reducing the unexplained variability or experimental error. If the covariates do indeed have an effect, this leads to more precise evaluations of the primary variables. Alternatively, such covariates may be related to the unexplained or residual variation that remains after the statistical analysis of the experimental results, using graphical or other techniques.

Stagewise Experimentation

Contrary to popular belief, a statistically planned experiment does not require all testing to be conducted at one time. Instead, such programs can, and often should, be conducted in stages of, eg, 8–20 runs; this permits changes to be made in later tests based on early results and allows preliminary findings to be reported. A recent experiment to improve the properties of a plastic material involved such variables as mold temperature, cylinder temperature, pressure, ram speed, and material aging. The experiment was conducted in three stages. After the first stage, the overall or main effects of each of the variables on the experimental responses were evaluated; after the second stage, interactions between pairs of variables were analyzed, and after the third stage nonlinear effects were assessed. Each stage involved about a month of elapsed time, and management was periodically apprised of progress. If unexpected results had been obtained at an early state, eg, poor results at one of the selected ram speeds, the later stages of the experiment might be changed.

Whether or not to conduct a particular experiment in stages depends on the program objectives and the specific experimental situation; a stagewise approach is recommended when units are made in groups or one at a time and a rapid feedback of results is possible. Running the experiment in stages is also attractive in searching for an optimum response, because it might permit a move closer to the optimum from stage to stage. On the other hand, a single-stage experiment may be desirable if there are large start-up costs at each stage or if there is a long waiting time between fabricating the units and measuring their performance. This is the case in many agricultural experiments and also when the measured variable is product life.

If the experiment is conducted in stages, precautions must be taken to ensure that possible differences between the stages do not invalidate the results. Appropriate procedures to compare the stages must be included, both in the test plan and in the statistical analysis. For example, some standard test conditions, known as controls, may be included in each stage of the experiment.

Other Considerations

Many other questions must be considered in planning the experiment:

- (1) What is the most meaningful way to express the controllable or independent variables? For example, should current density and time be taken as the experimental variables, or are time and the product of current density and time the real variables affecting response? Judicious selection of the independent variables often reduces or eliminates interactions between variables, thereby leading to a simpler experiment and analysis. Also inter-relationships among variables need be recognized. For example, in an atomic absorption analysis, there are four possible variables: air-flow rate, fuel-flow rate, gas-flow rate, and air fuel ratio, but there are really only two independent variables.
- (2) What is a proper experimental range for the selected quantitative controllable variables? Assuming a linear relationship between these variables and performance, the wider the range of conditions or settings, the better are the chances of detecting the effects of the variable. However, the wider the range, the less reasonable is the assumption of a linear relationship. Also, one generally would not want to conduct experiments appreciably beyond the range of physically or practically useful conditions. The selection of the range of the variables depends in part on the ultimate purpose of the experiment: is it to learn about performance over a broad region or to search for an optimum condition? A wider range of experimentation would be more appropriate in the first case than in the second.
- (3) What is a reasonable statistical model, or equation, to approximate the relationship between the independent variables and each response variable? Can the relationship be approximated by an equation involving linear terms for the quantitative independent variables and two-factor interaction terms only or is a more complex model, involving quadratic and perhaps even multifactor interaction terms, necessary? As indicated, a more sophisticated statistical model may be required to describe relationships adequately over a relatively large experimental range than over a limited range. A linear relationship may thus be appropriate over a narrow range, but not over a wide one. The more complex the assumed model, the more runs are usually required to estimate model terms.
- (4) What is the desired degree of precision of the final results and conclusions? The greater the desired precision, the larger is the required number of experimental runs.
- (5) Are there any previous benchmarks of performance? If so, it might be judicious to include these conditions in the experiment to compare the results.
- (6) What statistical techniques are required for the analysis of the resulting data, and can these tools be rapidly brought to bear after the experiment has been conducted?

Formal Experimental Plans

After the preceding considerations have been taken into account, a test plan is developed to best meet the goals of the program. This might involve one of the standard plans developed by statisticians. Such plans are described in various texts (Table 1) and are considered only briefly here. Sometimes, combinations of plans are encountered, such as a factorial experiment conducted in blocks or a central composite design using a fractional factorial base.

Table 1. Summary of Experimental Design Texts^a

Subject coverage													
Gener al	Blocki ng	Factori al	Fraction al-factori	Respon se-surfa	Split-pl ot	Nested situatio	Taguc hi	Num ber of	Emphasis on	Approxim ate	Exempl es	Tec h.	Applicatio ns

reference numbers	designs	designs	al designs	ce designs	situatio ns	ns	desig ns	pages	calculatio ns and analysis	number of references		leve l ^b	emphasis ^c
1	M-H	M-H	M-H	M-H	H	M-H	N	418	L-M	150	M	I	G I
2	M	M	L-M	M-H	N	L	N	653	L	150	M-H	E	G I
3	L-M	L-M	L-M	L	M	N	N	308	N	100	M	E	G
4	N-L	H	H	L	L-M	N	N	294	L-M	100	M	I	I
5	M-H	M-H	M-H	M-H	N	L-M	N	656	L-M	50	M	I	I
6	L	M	M-H	L	L-M	L	N	390	L-M	20	M	I	I
7	M	M-H	L-M	L-M	M	M	N	349	M	20	M	I	I
8	N	L	L	N	N	N	N	518	M	100	M	E	I
9	L-M	M	M	L	N	N	H	241	L	50	M-H	E	I
10	M	M-H	M	M	M	M-H	N-L	692	M-H	120	M	E-I	I
11	L-M	L-M	L-M	L-M	L	L-M	N	464	L-M	100	M	I	G I
12	M	M-H	M	L	L	M-H	L-M	414	L	30	M-H	E	I
13	M-H	M-H	M-H	M-H	M-H	M-H	M	649	M-H	120	L-M	E-I	I

^a N is none; L, little; M, moderate; and H, heavy; see General References for descriptions.
^b E is elementary, no background in statistics needed; and I is intermediate, assumes one or two elementary statistics courses.
^c G is general; I, industrial engineering scientific.

Blocking Designs. Such designs use blocking techniques to remove the effect of extraneous variables from experimental error. Well-known blocking designs include randomized and balanced incomplete block designs to remove the effects of a single extraneous variable, Latin square and Youden square designs to remove the effects of two extraneous variables, and Greco-Latin square and hyper-Latin square plans to remove the effects of three or more extraneous variables.

Complete Factorial and Fractional Factorial Designs. Such designs apply for two or more primary independent variables. Factors are varied simultaneously, rather than one at a time, so as to obtain information about interactions among variables and to obtain a maximum degree of precision in the resulting estimates. In complete factorial plans, all combinations of conditions of the independent variables are run. For example, a $3 \times 3 \times 2 \times 2$ factorial design requires running all 36 combinations of four variables at three, three, two, and two conditions respectively. A fractional factorial design is often used when there is a large number of combinations of possible test points, arising from many variables or many conditions per variable or both, and it is not possible or practical to run all combinations. Instead, a specially selected fraction is run. For example, a $(1/2)^2$ fractional factorial plan is one where there are six variables each at two conditions, resulting in a total of 64 possible combinations, but only a specially selected one-half, or 32, of these combinations are actually run. Fractional factorial plans are especially useful for screening purposes when it is desired to find those variables that have the greatest impact (over the specified experimental region). Reference 13 provides a comprehensive catalogue of fractional factorial designs.

An additional advantage of full factorial and fractional factorial designs is that by providing a comprehensive scanning of the experimental region they can often identify, without any further analyses, one or two test conditions that are better than any others. The region around these conditions can then be explored further in subsequent experimentation.

Response Surface Designs. These are special multivariable designs for quantitative independent variables (temperature, pressure, etc). The relationship between the independent variables and the response variable is fitted using least-squares regression analysis. Popular response surface designs include the orthogonal central composite (or Box) designs, rotatable designs, and Box-Behnken designs.

Mixture Designs. These designs arise when the variables under consideration are the ingredients of a product that must add to 100%, eg, in an experiment to learn how to optimize the baking of a cake, the relative impact of each of the ingredients (baking powder, shortening, flour, sugar, milk, water, etc) might need to be assessed. Most standard texts do not discuss mixture experiments. However, detailed expositions are provided in References 14 and 15.

Taguchi Designs. An eminent Japanese engineering professor and advocate for quality improvement, Genichi Taguchi has been an influential proponent of planned experimentation in product design, development, and evaluation. He has helped popularize "orthogonal arrays," a family of experimental designs closely related to fractional factorial plans. Taguchi has also stressed the role of experimental design in identifying the process conditions that show the greatest consistency, or are most robust, in face of variability in manufacturing conditions and customer use. The goal is to achieve similar product performance despite uncontrollable variation in, say, incoming raw materials, and the ambient conditions under which a product is to be used. For example, an automobile battery needs to work well, irrespective of whether a car is driven in Florida or Minnesota. Taguchi has also proposed various methods of data analysis and other concepts, some of which are controversial. A series of articles that describe, illustrate, and critique Taguchi methods is provided in Reference 16.

Computer-Aided Design of Experiments

In recent years, numerous computer software packages have been developed to help generate experimental designs (17). These packages vary greatly in price, computer platform, and level of technical sophistication assumed of the user. They are especially useful in performing the mechanical tasks associated with the design and analysis of experiments, such as the generation of a matrix of test points and, in some cases, the statistical analysis of the resulting data. They can also be helpful in generating "statistically optimum" test plans for special situations, such as plans to void regions of little practical interest, or conditions that might be hazardous to run (see COMPUTER AIDED ENGINEERING).

Such programs generally concentrate on the technical parts of designing an experiment, and provide limited guidance on the important, softer aspects of experimental design stressed in this article. Also, most computer routines do not allow one to handle various advanced concepts that arise frequently in practice, eg, split plot and nested situations, discussed in the books in the bibliography. In fact, some of the most successful experiments do not involve standard canned plans, but are tailored to fit the problem at hand.

BIBLIOGRAPHY

"Design of Experiments" in *ECT* 3rd ed., Vol. 7, pp. 526–538, by G. J. Hahn, General Electric Co.

- G. J. Hahn, *Chemtech* **5**, 496 (Aug. 1975); **5**, 561 (Sept. 1975).
- G. J. Hahn, *J. Quality Technol.* **9**, 13 (Jan. 1977).
- D. E. Coleman and D. C. Montgomery, *Technometrics* **35**, 1 (Feb. 1993), with discussion.
- F. X. Mueller and D. M. Olsson, *J. Paint Technol.* **43**, 54 (May 1971).
- G. J. Hahn, *Technometrics* **26**, 19 (Feb. 1984).
- C. D. Hendrix, *Chemtech* **16**, 228 (1986).
- J. S. Hunter, *Chemtech* **17**, 187 (1987).
- J. S. Hunter, in J. M. Juran, ed., *Quality Control Handbook*, 4th ed., McGraw-Hill Book Co., New York, 1988, Chapt. 26.

9.

F. G. Tingey, in I. M. Kolthoff and P. J. Elving, eds., *Treatise on Analytical Chemistry*, Pt. 1, Wiley-Interscience, New York, 1972, Chapt. 106.

10.

D. M. Steinberg and W. G. Hunter, *Technometrics* **26**, 71 (May 1984), with discussion.

11.

G. J. Hahn, *Chem. Technol.* **7**, 630 (Oct. 1977).

12.

G. J. Hahn, *Chem. Technol.* **8**, 164 (Mar. 1978).

13.

G. J. Hahn and S. S. Shapiro, *A Catalog and Computer Program for the Design and Analysis of Orthogonal Symmetric and Asymmetric Fractional Factorial Experiments*, General Electric Co. TIS Report 66C165, General Electric Co., Schenectady, N.Y., May 1966.

14.

R. D. Snee, *J. Quality Technol.* **3**, 159 (Oct. 1971).

15.

J. A. Cornell, *Experiments with Mixtures*, John Wiley & Sons, Inc., New York, 1990.

16.

Qual. Relia. Eng. Int. **4**, 193 (Aug. 1988).

17.

G. J. Hahn and C. B. Morgan, *Chemtech* **18**, 664 (Nov. 1988).

General References

Described below are books that deal mainly with the application of experimental design to scientific, industrial, and general situations. Books directed principally at educational, psychological, or related applications, and those dealing mostly with the theory of experimental design, or exclusively with the analysis of experimental data, are omitted. Most of the books presume that the reader has had at least one introductory course in statistics. Items in quotation marks are taken directly from the book, generally from its preface. The book by Box, Hunter, and Hunter (1978) is, perhaps, the one that has become best known among practitioners. Table 1 provides a comprehensive summary of the subject coverage, technical level, applications emphasis, etc of these books. In addition, such journals as *CHEMTECH*, the *Journal of Quality Technology*, and *Technometrics* (in increasing order of complexity) regularly carry articles about experimental design.

V. L. Anderson and R. A. McLean, *Design of Experiments—A Realistic Approach*, Marcel Dekker, New York, 1974. This book provides an extensive exposition of experimental design at a relatively elementary level. It includes most of the standard material, as well as detailed discussions of such subjects as nested and split-plot experiments. Restrictions on randomization receive special emphasis.

G. E. P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, John Wiley & Sons, Inc., New York, 1978. This book, by three eminent practitioners, "is an introduction to the philosophy of experimentation and the part that statistics plays in experimentation." It provides a practically motivated introduction to basic concepts and methods of experimental design. It "is written for those who collect data and try to make sense of it," and gives an "introduction to those ideas and techniques" that the authors have found especially useful. Statistical theory is introduced as it becomes necessary. Readers are assumed to have no previous knowledge of the subject; "the mathematics needed is elementary." The book includes numerous examples and case studies and provides appreciable detail about many elementary and a few more advanced designs. It is, however, an introductory treatment; therefore, more advanced situations are not discussed. Heavy emphasis is placed on the use of graphical methods for analyzing experimental results. In contrast, the more computationally involved analysis-of-variance tools receive relatively little attention. The four parts of the book deal with (1) comparing two treatments, (2) comparing more than two treatments, (3) measuring the effects of variables, and (4) building models and using them.

D. R. Cox, *Planning of Experiments*, John Wiley & Sons, Inc., New York, 1958. This book provides a simple survey of the principles of experimental design and of some of the most useful experimental schemes. It tries "as far as possible, to avoid statistical and mathematical technicalities and to concentrate on a treatment that will be intuitively acceptable to the experimental worker, for whom the book is primarily intended." As a result, the book emphasizes basic concepts rather than calculations or technical details. Chapters are devoted to such topics as "Some key assumptions," "Randomization," and "Choice of units, treatments, and observations."

C. Daniel, *Applications of Statistics to Industrial Experimentation*, John Wiley & Sons, Inc., New York, 1976. This book is based on the personal experiences and insights of the author, an eminent practitioner of industrial applications of experimental design. It provides extensive discussions and concepts, especially in the areas of factorial and fractional factorial designs. "The book should be of use to experimenters who have some knowledge of elementary statistics and to statisticians who want simple explanations, detailed examples, and a documentation of the variety of outcomes that may be encountered." Some of the unusual features are chapters on "Sequences of fractional replicates" and "Trend-robust plans," and sections entitled, "What is the answer? (what is the question?)," and "Conclusions and apologies."

O. L. Davies and co-workers, *The Design and Analysis of Industrial Experiments*, 2nd ed., Hafner, New York, 1956; reprinted by Longman, New York, 1987. This book, which is a sequel to the authors' basic text *Statistical Methods in Research and Production*, is directed at industrial situations and chemical applications. Three chapters are devoted to factorial experiments and one chapter to fractional factorial plans. A lengthy chapter (84 pp.) discusses the determination of optimum conditions and response surface designs, which are associated with the name of G. Box, one of the seven co-authors. Theoretical material is presented in chapter appendices.

W. J. Diamond, *Practical Experimental Designs for Engineers and Scientists*, 2nd ed., Van Nostrand Reinhold, New York, 1989. "This book is for engineers and scientists with little or no statistical background who want to learn to design efficient experiments and to analyze data correctly ... The emphasis is on practical methods, rather than on statistical theory." The discussion is quite detailed in some areas, eg, experimental designs based on Hadamard matrices, and scanty in others.

C. R. Hicks, *Fundamental Concepts of Design of Experiments*, 3rd ed., Holt, Rinehart and Winston, New York, 1983. "It is the primary purpose of this book to present the fundamental concepts in the design of experiments using simple numerical problems, many from actual research work ... The book is written for anyone engaged in experimental work who has a good background in statistical inference. It will be most profitable reading to those with a background in statistical methods including analysis of variance." This work provides an intermediate coverage of most basic experimental designs.

C. Lipson and N. J. Sheth, *Statistical Design and Analysis of Engineering Experiments*, McGraw-Hill, New York, 1972. "This book is written in a relatively simple style so that a reader with a moderate knowledge of mathematics may follow the subject matter. No prior knowledge of statistics is necessary." Appreciably more discussion is devoted to statistical analysis than to the planning of experiments. Some relatively nonstandard subjects (for an introductory text), such as accelerated experiments, fatigue experiments, and renewal analysis are also included.

R. H. Lochner and J. E. Matar, *Designing for Quality: An Introduction to the Best of Taguchi and Western Methods of Statistical Experimental Design*, Quality Resources, White Plains, N.Y., and ASQC Quality Press, Milwaukee, Wis., 1990. "The purpose of this book is to show engineers with little or no previous exposure to experimental design how to use statistically designed experiments to improve products and processes. A prerequisite ... is an appreciation for the concept of random variation, as is gained through a typical statistics course or by using Statistical Process Control (SPC) in a manufacturing environment." As its subtitle suggests, this book merges a discussion of concepts introduced or popularized by G. Taguchi, with those of classical experimental design.

R. L. Mason, R. F. Gunst, and J. L. Hess, *Statistical Design and Analysis of Experiments with Applications to Engineering and Science*, John Wiley & Sons, Inc., New York, 1989. "The intended audience of this book consists of two groups. The first group covers a broad spectrum of practicing engineers and scientists, including those in supervisory positions ... This audience includes those who have little formal training in statistics ... The second group ... is students in introductory statistics courses ... in which statistical experimental design and the analysis of data are the main topics ... College algebra is the only prerequisite." This book, in addition to discussing the design of experiments, provides a comprehensive introduction to modern statistical methods. Following two introductory chapters, the four parts of this book deal with describing variability, experimental design, analysis of designed experiments, and fitting data.

W. Mendenhall, *Introduction to Linear Models and the Design and Analysis of Experiments*, Duxbury Press, Belmont, Calif., 1968. This book provides an introduction to basic concepts and the most popular experimental designs without going into extensive detail. In contrast to most other books, the emphasis in the development of many of the underlying models and analysis methods is on a regression, rather than an analysis-of-variance, viewpoint.

R. D. Moen, T. W. Nolan, and L. P. Provost, *Improving Quality through Planned Experimentation*, McGraw-Hill, New York, 1991. "This book is about planned experimentation to improve quality." It aims to promote "a substantial increase in the number of people who will use these methods" by requiring "a lower

level of mathematical and statistical sophistication than was previously necessary ... Taguchi's contributions to the application of experimentation to the design of products and processes have been integrated ... (and) the book is written to be compatible with (W. E. Deming's viewpoint of analytic studies." Graphical, rather than formal statistical, methods for data analysis are presented.

D. C. Montgomery, *Design and Analysis of Experiments*, 3rd ed., John Wiley & Sons, Inc., New York, 1991. This "introductory textbook dealing with the design and analysis of experiments ... is intended for readers who have completed a first course in statistical methods." It provides a basic treatment of standard experimental plans and techniques for analyzing the resulting data. Originally published in 1976, the revised edition includes an introduction to mixture experiments and Taguchi designs.

Gerald J. Hahn
General Electric Company

DETERGENCY

The cleaning of a solid object, ie, the removal of unwanted foreign matter from its surface, is done by methods ranging from simple mechanical separation such as blotting or abrasion to removal by solution or selective chemical action. The term detergency is limited to systems in which a liquid bath is present and is the main cleaning component of the system. The action of the bath involves more than simple solution or simple hydraulic dislodging of soil, although both will occur and contribute to the cleaning. The cleaning is enhanced primarily by the presence in the bath of a special solute, the surfactant, which alters interfacial effects at the various phase boundaries within the system. Thus, a typical deterative system consists of a solid object to be cleaned, called the substrate; soil or dirt attached to it which is to be removed in the washing process; and a liquid bath that is applied to the soiled substrate. In turn, each of these elements can vary widely in properties and composition. The final cleaning benefit results from interaction of these elements and the conditions used, ie, temperature, time, mechanical energy input (agitation) and, in the case of aqueous baths, the presence of hardness ions in the water.

In the cleaning or washing process in a typical deterative system the soiled substrate is immersed in or brought into contact with a large excess of the bath liquor. Enough bath is used to provide a thick layer over the whole surface of the substrate. During this stage, air is displaced from soil and substrate surfaces, ie, they are wetted by the bath. The system is subjected to mechanical agitation, either rubbing or shaking, which provides the necessary shearing action to separate the soil from substrate and disperse it in the bath. Agitation also promotes mass-transfer in the system, just as in a heterogeneous chemical reaction. The bath carrying the removed soil is drained, wiped, squeezed, or otherwise removed from the substrate. The substrate is rinsed free of the remaining soiled bath. This rinsing step determines the final cleanliness of the substrate. The cleaned substrate is dried or otherwise finished.

A meaningful discussion of detergency requires a definition of clean. In the physiochemical sense, a surface is clean if it contains no molecular species other than those in the interior of the two adjoining phases. It is difficult to achieve such a state even under the most exacting laboratory conditions. Practically, a surface is clean if it has been brought to a desired state with regard to foreign matter present upon it, as judged by agreed upon criteria. Household linen, for example, is considered clean when it is free of visible soil even though it may carry a starch and a softening finish. In the dyehouse of a textile mill, a piece of goods such as this would be rejected as dirty and returned for scouring because these finishes interfere with dyeing. Most standards for cleanness involve a visual or optical judgment for the presence of foreign matter. In some systems, for example, the desizing of cotton, the degree of cleanness may be specified by weight percentage of soil on the substrate. In other systems, such as the degreasing of metal, it is the weight of soil per unit area of substrate surface that specifies cleanness. In washing dishes or glassware, cleanness is often specified by complete water wettability or freedom from water break, as well as by appearance (see also METAL SURFACE TREATMENT; TEXTILES).

Although it is impossible to list all the practical deterative systems that might be encountered, a large proportion fall in a small number of classes. This classification disregards surfactant structure and type of substrate (fibrous or hard surface) and is restricted to a consideration of the soil present on the substrate, the mechanical action employed, the bath ratio, and the detergent used. Some of the more commonly encountered deterative systems are classified on this basis in Table 1.

Table 1. Deterative Systems

System	Soil	Mechanical action	Bath ratio	Detergent
<i>Fabric and fibrous</i>				
textile and allied manufacturing				
raw wool scouring	liquid at operating temperature	very gentle	high	organic surfactant
wool yarn and piece goods	liquid	gentle	moderately high	organic surfactant
scouring gray cottons	mostly solid; waxy and starchy	vigorous	low	built surfactant
scouring synthetic fabrics	mixed; oily lubricant and sizing residues	moderate	medium	built surfactant
de-inking paper	mixed; oily and pigment	vigorous	high	heavily built surfactant
laundering				
household	mixed and variable	moderate	high	unbuilt or built surfactant
commercial and industrial	mixed and variable; heavier than household soil	vigorous	high	built surfactant
rug cleaning in plant	heavy solid	moderate gentle	medium to low	organic surfactant, may be built
on location	heavy solid	vigorous; superficial brushing	very low; foam bath	organic surfactant foam
<i>Hard surfaces</i>				
glass and ceramics				
hand dishwashing	mixed oily and solid organic	moderate to vigorous	high	organic surfactant
machine dishwashing	mixed oily and solid	vigorous hydraulic	high	inorganic ^a
commercial bottle washing	light solid organic	vigorous hydraulic	high	inorganic ^a
metals				
prefinishing cleaning	oily	moderate to gentle hydraulic	high	inorganic ^a
postforming emulsion cleaning	oily and mixed	gentle hydraulic	high	mixed inorganic and surfactant
cleaning metal structures and equipment, tanks, dairy equipment, etc	variable; mostly oily or organic solid	usually vigorous rubbing, sometimes hydraulic	usually low in wash cycle; may be high in rinse cycle	inorganic ^a , built surfactant
organic surfaces, paint, linoleum, plastic tile	mixed solid, organic solid, and oily	usually vigorous low rubbing	low	lightly to moderately built surfactant
<i>Cosmetic and personal cleaning</i>				
shampooing	oily	vigorous rubbing in rinse cycle	low in wash cycle; high	organic surfactant
bathing and washing	mixed, mostly oily	mild to vigorous rubbing	high; sometimes low in wash cycle	organic surfactant

^a Organic surfactant frequently added to aid wetting and draining.

Components of Detersive Systems

Substrates. Solid objects to be cleaned vary widely in chemical composition and surface configuration. With few exceptions, however, they can be divided into fabrics and fibrous materials, and hard surfaces. Fabrics present a highly complex configuration and can entrap soil even though it may be physicochemically removed from the surface. Most fabrics are organic polymeric materials that may be swellable by water or permeable to small molecules and ions dissolved in the bath. Many common polymeric fabrics, cotton, polyester, rayon, nylon, wool and cellulose esters, contain ionogenic or polar centers capable of localizing (generally anionic) electric charges. Hard surfaces, on the other hand, are relatively flat and smooth. They cannot entrap soil that has been detached by physicochemical action. In general, they are impermeable to water and water-soluble materials although they vary widely in their wettability and polarity. The most important types are glass, metal, and organic polymeric materials including painted surfaces, linoleum, and plastic tile.

Soils. Soils vary greatly. They may be a single solid or liquid phase but usually are two or more phases, intimately and randomly mixed and irregularly disposed over the substrate. In a large number of important detersive systems, the nature of soil and the quantity present are well known. This is the case, for example, in most textile mill operations such as raw wool scouring, the boil-off and scouring of loom-state woven goods, and the soaping of printed cottons. In the cleaning of fabricated metal parts to remove forming and drawing lubricants, buffing compounds, etc, the nature of the soil is well known. The behavior of soils encountered in dishwashing is well characterized.

As a result of many painstaking investigations, the soils on apparel encountered in laundering have been shown to be complex mixtures containing both oily and finely divided solid material (1,2). The oily material consists largely of fatty acids and polar fatty material but a considerable proportion of neutral nonpolar oil is also present. The solid components vary widely with the locale in which samples are taken, and resemble local street dust in composition.

Particle size is one of the most important factors in determining the ease with which solid soil can be removed from a substrate. Particles of >5 μm dia are generally easily removed. Particles of <10 nm dia cannot be removed by ordinary detersive processes once they are attached to a typical textile fabric. Such particles are responsible for the gradual irreversible graying of white goods with continued wear and laundering. Particles in this size range tend to form clumps and clusters before they reach the fiber surface. These clusters behave like individual large particles. Particles or clusters in the range of 100 nm dia resist removal by simple agitation in liquids that are not surface active, but these particles are removable by normal detersive processes. This is the size range of greatest interest.

Soil may include material that is soluble in the bath, such as encrusted sugar residues and molecularly dispersed material such as fruit juice stains. Removal of these soils is an important aspect of cleaning but is not generally considered in discussions of detergency.

Baths. The baths discussed here are aqueous solutions. Most nonaqueous cleaning systems, such as metal degreasing operations, depend entirely on solvent action and therefore cannot be considered examples of detersive systems. Some nonaqueous systems, however, are true detersive systems. Modern dry cleaning baths, for example, contain solutes that are surface active in the conventional hydrocarbon or chlorinated hydrocarbon medium and aid soil removal. The physical chemistry of such systems differs considerably from that of aqueous systems. Among bath components the solute that is effective in cleaning, usually a mixture of several components, is called the detergent. The term detergent is also used frequently in the restricted sense of a surfactant of high detersive power. In many hard-surface systems, however, nonsurfactants such as alkaline silicates and phosphates exert a true detersive effect. They are, in fact, the principal detergents in these systems even in the complete absence of any surfactants. In the cleaning of fabric systems, the most important detersive component in the bath is the surfactant. Nonsurfactant components that augment the cleaning effect of surfactants are called builders. Many materials that act as builders in fabric systems, eg, phosphates and silicates, are the primary detergents in hard-surface systems, although their primary contribution to the cleaning process may differ in the two cases.

Formulation

Detergents are formulated to clean a defined set of soiled substrates under an expected range of washing conditions. Some detergents, the familiar bar or toilet soap, for example, consist essentially of only one component. There are few systems, however, in which a suitably formulated detergent consisting of several components does not outperform the best single-component system. Although detergents for hand dishwashing rarely contain builders, those currently used in the U.S. contain at least three surfactants, and may contain up to six. Ingredients of laundering detergent formulations for fabrics may be divided into the following groups: surfactants (qv), including soap and various others; the inorganic salts, acids, and bases, including builders, and other compounds that do not contribute to detergency but provide other functions, such as regulating density and assuring crispness of powdered formulations; organic additives that enhance detergency, foaming power, emulsifying power, or soil-suspending effect of the composition; and special purpose additives, such as bleaching agents (qv), fluorescent whitening agents (qv), antimicrobial agents, blueing agents, or starch, which provide desirable performance functions but have no direct effect on soil removal (see also INDUSTRIAL ANTIMICROBIAL AGENTS).

Fabric detergent formulations for special applications, such as the various specific operations within the textile mill, are frequently much simpler. They tend to contain little if any builder or special-purpose additive. The indispensable ingredient in fabric detergency is the organic surfactant. Formulations for hard-surface detergency such as those used in automatic machine dishwashing, are simpler than fabric-washing compositions. An organic surfactant is frequently not needed and inorganic salts are the detersive ingredients.

Surfactants. The most important components of detersive systems are, of course, the surfactants described elsewhere in the *Encyclopedia*.

Builders. Builders are substances that augment the detersive effects of surfactants (3). Most important is the ability to remove hardness ions from the wash liquor (ie, soften the water) and thus to prevent them from interacting with the surfactant. Such interaction reduces detersive effectiveness. Hardness ions can also interact with the negative charges present on soil and fabric surfaces (formed, eg, by ionization of –OH or –COOH groups) reducing electric repulsion between them, thus hindering the detersive process. Hardness ions are best removed by sequestration to form soluble chelates (see CHELATING AGENTS). Less desirable is the formation of insoluble precipitates that may deposit on fabrics and machines and can, over many wash-and-wear cycles, lead to incrustation on machine parts and harshening of fabric. A third mechanism for removing hardness ions from wash liquors is through ion exchange in which calcium in solution displaces sodium ions in the ion exchanger, thus effectively removing the hardness from solution. The ion exchanger, in general, is a solid. Unlike precipitated calcium carbonate, however, the particle size of the ion exchanger can be controlled and the problem of the presence of insoluble particles in the wash liquor can be reduced.

In general, builders supply alkalinity to the wash liquor and thus function also as alkalies. In addition, they can exert a suspending (antiredeposition) effect and keep detached soil from depositing on the fabric; builder ions with multiple charges are especially effective in this area.

Phosphates. Pentasodium triphosphate [7758-29-4], sodium tripolyphosphate, STPP, Na₅P₃O₁₀, is the most widely used and most effective builder in heavy-duty fabric washing compositions (see also PHOSPHORIC ACID AND PHOSPHATES). It is a strong sequestrant for calcium and magnesium, with a pK^{Ca} of ca 6, and provides excellent suspending action for soils. Because of its high sequestration power, it also finds extensive application in automatic-dishwashing detergents. Sodium tripolyphosphate forms stable hydrates and thus aids in the manufacture of crisp spray-dried laundry powders.

Tetrasodium pyrophosphate [7722-88-5], Na₄P₂O₇, is another important primary builder and detergent. In sequestration, it is not quite as effective as sodium tripolyphosphate and its usage in heavy-duty laundry powders has declined in recent years. Functionally, tetrasodium pyrophosphate is both a builder for surfactants (ie, water softener) and alkali.

Where hardness is present in excess of the sequestering capacity of sodium tripolyphosphate and pyrophosphate, both can function as precipitant builders.

Trisodium phosphate [7601-54-9], trisodium orthophosphate, Na₃PO₄, is an important constituent of hard-surface cleaners including those for ceramic, metal, or painted surfaces. It may be used with soaps, surfactants, or other alkalies. It precipitates many heavy-metal ions but does not sequester to form soluble chelates. It is thus a precipitant builder and additionally an alkali.

Glassy phosphates (sodium polymetaphosphate [50813-16-6], sodium hexametaphosphate [10124-56-8]) vary in composition, depending upon the

manufacturing process. They exert a powerful sequestering and suspending effect combined with a low solution pH, about 6 or 7, and tend to hydrolyze or revert in aqueous solution and heat to pyrophosphates and orthophosphates.

Potassium phosphates, particularly tetrapotassium pyrophosphate [7320-34-5], $K_4P_2O_7$, are considerably more soluble than their sodium analogues. They have been used as builders in liquid detergents.

In the early 1970s, a number of U.S. jurisdictions banned the use of phosphates as detergent builders. The trend has continued to the point today that phosphates are banned in a sizeable proportion of the U.S. (about 45%^o). Consequently, several alternatives to phosphates have been introduced into heavy-duty U.S. laundry detergents. No entirely satisfactory single substitute for sodium tripolyphosphate has been found that is as cost effective. Sodium tripolyphosphate aids detergency not only via water softening (calcium–magnesium sequestration) but also via soil suspension, soil removal, and antiredeposition benefits, all of which are closely related mechanistically. Additionally, STPP provides excellent spray-dried powder properties. However, it has been found that use of a water softener such as 4A Zeolite, in combination with a mixed active system, buffer such as carbonate, and soil suspension antiredeposition agents such as NaCMC, poly(ethylene glycol), polacrylate, polyacrylate–maleate copolymers, plus other cobuilders such as citrate, can provide general cleaning at least as good as the old high P formulations. Indeed current products containing enzymes and effective low temperature bleach are superior. It seems likely that all principal U.S. detergent manufacturers will remove phosphate from their products in the near future.

Sodium Carbonate. Sodium carbonate softens water by forming insoluble calcium carbonate with calcium ions in hard water. Carbonate can also reduce calcium levels by ion pairing, although the benefit to detergency is questionable. Buildup of calcium carbonate on machine and fabrics, which can occur with time, is undesirable. Sodium carbonate [497-19-8] does not provide any suspending action. It does, however, provide alkalinity to the wash liquor and is an effective alkali.

Silicates. Sodium silicates have been used extensively as soap builders in laundering formulations since well before the advent of synthetic surfactants. Silicates are more effective in removing magnesium than calcium hardness. Again, they function primarily as alkalies. In addition, they act as anticorrosive agents and prevent deterioration of washing machines, specifically metal pump parts. However, in recent years, many of these machine parts have been replaced by engineering plastics and the anticorrosion function has lost some of its importance. Alkaline silicates act as primary detergents in machine-dishwashing formulations. Commercial alkaline silicates are characterized by the ratio of SiO_2 to Na_2O in the molecule. The silicates used in detergent formulations generally show a ratio >1, usually 2.0 to 2.4. The 1:1 compound, sodium metasilicate, is considered too corrosive to be widely used in consumer product formulations.

Zeolites. Certain zeolites have found application as builders in heavy-duty detergent formulations. The zeolite of choice is a so-called Type A zeolite, of empirical composition $Na_2O \cdot Al_2O_3 \cdot 2SiO_2 \cdot 4.5H_2O$ and particle size of the range of 10 μm . This builder functions by ion exchange in which sodium ions released from the zeolite crystal are replaced by calcium ions in hard water, thus lowering the free hardness in the wash. Its pore size accommodates calcium ions but is not sufficiently large for the highly hydrated magnesium ions. Zeolite A, therefore, is not an effective builder for magnesium hardness. Like sodium carbonate building, the ion exchange process is appreciably slower than soluble chelate formation by strong sequestrants like sodium tripolyphosphate. Type A zeolite is used principally to replace sodium tripolyphosphate in areas where phosphates are limited by law, ie, in the United States, Canada, and Western Europe. However, it is not practical as the sole builder in a nonphosphate detergent formulation (see also MOLECULAR SIEVES), since it does not contribute to alkalinity, soil suspension, or bind magnesium.

Clays. Clays (qv), such as kaolin, the montmorillonites, and bentonites, have been recommended and used from time to time as ingredients of washing compositions and other formulations containing surface-active agents. Under certain favorable conditions, particularly in soft water of low dissolved solids content, clay suspensions can have a marked deterative effect on ordinary soiled fabrics. Bentonite [1302-78-9] also acts as a suspending agent. In addition, sodium bentonite has some water softening effectiveness by virtue of its ability to sorb calcium ions. However, clays are considerably less effective than Type A zeolite in water softening.

In one U.S. laundry powder, a montmorillonite clay serves as the main softening component. It is combined with a waxlike cationic granule (4–7). Both are absorbed or filtered onto the cloth during the wash and spin rinse. The clay absorbs onto the fabric in thin sheetlike layers, providing a lubricating effect. The cationic particles melt in the heat of the automatic dryer providing an antistatic benefit and augmenting the softening benefit of the clay. A softening-in-the-wash effect is thus achieved with minimal interference with detergent performance.

Nitriiotriacetic Acid. The trisodium salt [5064-31-3] of nitriiotriacetic acid, $N(CH_2COOH)_3$, so-called NTA, is a powerful sequestrant builder, comparable to sodium tripolyphosphate. It is therefore noted here, even though it is an organic builder. NTA has been recommended and used as a phosphate replacement in areas where phosphate is banned. However, because of adverse laboratory reports of possible teratogenic effects, NTA has withdrawn in 1970 from consumer products at the suggestion of the U.S. Surgeon General. Because it is a smaller molecule than sodium tripolyphosphate, NTA is theoretically a more effective sequestrant on a weight basis. It is, however, less effective than sodium tripolyphosphate as a suspending agent and is not as easily processed in spray-dried laundry powders.

Alkalies. Caustic soda (sodium hydroxide) is used largely in mechanical bottle washing, glass washing, and metal cleaning. Sodium carbonate, either anhydrous (soda ash) or in hydrated form, has been used as builder or filler in soaps, surfactants, and with inorganic constituents in cleaners for hard surfaces and fabrics. It forms insoluble calcium carbonate with calcium ions in hard water but does not provide any suspending action. Sodium bicarbonate [144-55-8]/ $NaHCO_3$, sodium sesquicarbonate [6106-20-3]/ $NaHCO_3 \cdot Na_2CO_3$, and sodium borate [1330-43-4]/ $Na_2B_4O_7 \cdot 10H_2O$, borax [1303-96-4] are used in place of soda ash when a lower pH is desired. In cases where high solubility is required, the potassium analogues are used.

The alkalies do not sequester heavy-metal ions and have little soil-suspending effect. They are effective in maintaining a high pH and saponify the acidic constituents of soil and thus promote cleaning. In the cleaning of ceramics, glass, and metal surfaces, however, the alkalies act as primary detergents even in the absence of surfactants in these systems.

Neutral Soluble Salts. Sodium sulfate [7757-82-6] and, to a considerably lesser extent, sodium chloride [7647-14-5] are the principal neutral soluble salts used in laundering compositions. They are often considered to be fillers although they perform an important standardizing function enabling the formulator to manufacture powders of a desired, controlled density. Sodium sulfate, in addition, lowers the critical micelle concentration of organic surfactants and thus the concentration at which effective washing can be achieved.

In wool-scouring systems for textile processing that contain nonionic surfactants, sodium chloride acts as a true builder, ie, detergency promoter.

Organic Additives. Certain nonsurfactant organic additives improve cleaning performance and exhibit other desirable properties.

Such additives are usually present in low percentage and serve one or more of the following specific functions: reduced redeposition of soil from the detergentbath onto the substrate; increased whiteness or appearance of cleanliness; enhanced cleaning effect on specific types of solid and stains; promotion or inhibition of foaming power and stability; increased solubility or other modification of the physical form of the detergent composition; sequestering of heavy-metal ions, both in the concentrated detergent and in the diluted cleansing bath; reduced injurious effects the detergent may have on the substrate or the washing machine, such as tarnishing of silverware or etching of glassware, corrosion of metals, or irritation of skin in toilet and cosmetic applications.

Antiredeposition agents contribute to the appearance of washed fabrics. Sodium carboxymethylcellulose [9004-32-4], NaCMC is the most widely used, and on cotton fabrics, the most effective. With the advent of synthetic fabrics, other cellulose derivatives, eg, methylcellulose [9004-67-5], hydroxybutylcellulose, hydroxypropyl- and mixed methyl and hydroxybutylcellulose ethers have been shown to be more effective than NaCMC (8) (see CELLULOSE ETHERS).

Fluorescent whitening agents (qv) were first disclosed in 1940 in combination with detergents (9). They absorb ultraviolet radiation and subsequently emit some of the radiation energy in the blue part of visible spectrum. As a result, they confer enhanced whiteness to the appearance of washed articles. Highly effective cotton-substantive fluorescent whitening agents were in widespread use at relatively high concentrations (ca 0.5%^o) in the 1950s and 1960s. For synthetic fabrics such as polyester, it has proved to be more effective to prebrighten the fabric by incorporating the fluorescent whitening agent in the spin-melt during manufacture rather than depend on adsorption from the detergent bath. As a result, the usage of fluorescent whitening agents in formulated laundry products has decreased in recent years.

Blueing agents, which are dyes, provide another approach to maintaining fabric whiteness by a mechanism in which a yellow cast of washed fabrics is covered by the blue dye. Since this approach reduces reflectance, it is less desirable than the use of fluorescent whitening agents that increase reflectance.

Proteolytic enzymes have been generally used in European detergent formulations. In the U.S., they were used in the late 1960s and early 1970s, but their use declined until they were present in only a few detergents. There has been a resurgence of their use in the last decade. Proteolytic enzymes in particular are widely used in premium products. They degrade proteinaceous stains and aid the cleaning performance of other formulation ingredients (10–12). Amylases and lipases have been used in a few U.S. detergents, the former to remove starches and the latter fatty esters and triglycerides. Cellulases have appeared in a few laundry detergents around the world. Since there are few, if any, cellulase-based soils present on home laundry, any laundering benefit from cellulase would be expected to come from action on cotton fabric. The nature and magnitude of such benefits is uncertain (see ENZYMES, INDUSTRIAL APPLICATIONS).

Bleaching agents (qv), such as sodium perborate trihydrate [28962-65-4], NaBO₃·3H₂O, are commonly present in significant amounts in European laundry detergents. At high washing temperatures, sodium perborate effectively bleaches chemical stains such as wine, fruit juices, etc. As European wash temperatures have declined so has the efficacy of perborate alone. Bleach activators, primarily TAED (tetracetyl ethylenediamine) have been widely used in Europe to provide effective bleaching at these low temperatures. Because of the lower washing temperature in U.S. machines, sodium perborate is considerably less effective and its usage is restricted to some individual brands of laundry detergents (13,14). In the U.S., even TAED is ineffective due to still lower wash temperatures, shorter wash times, and lower product concentration. One U.S. detergent manufacturer has introduced detergents with the bleach activator sodium nonanoyloxybenzene sulfonate (15). This activator forms the surface active species pernonanoic acid that does provide a bleach benefit under U.S. conditions. In automatic-dishwashing formulations, bleaching agents are needed to remove food stains from dishware and break down proteinaceous soil. Chlorine is the most cost-effective agent available for this purpose and is present in all U.S. products as chlorinated isocyanurate (see CYANURIC AND BIOCYNURIC ACIDS).

Foam regulators such as amine oxides, alkanolamides, and betaines are present in products where high foam value is functionally or esthetically desirable, mainly hand-dishwashing liquids and shampoos. In automatic dishwashing products, on the other hand, copious foam volumes interfere with the efficiency of the mechanical rotors during operation. In this type of product, a foam depressant is often present.

Organic sequestering agents serve the same purpose as the sequestering phosphates, ie, to remove interfering metal ions from the detergent bath. They would appear to also provide some benefits through ionic strength and soil suspending effects. They are used where the less expensive phosphates are, for one reason or another, not applicable. Nitrilotriacetic acid, EDTA, and a variety of organic phosphonate structures are commonly used in a wide range of detergent compositions. With the elimination of phosphate, sodium citrate [68-04-2] and synthetic analogues such as sodium tartrate monosuccinate and sodium tartrate disuccinate (16) have found wide use, especially in heavy-duty liquids. They are used to some extent in detergent powders. In the latter case they can also be used to modify powder properties, as can sodium polyacrylates. Certain hard-surface cleaning products such as sanitizing cleaners for hospital use are examples of products generally containing small amounts of organic sequestrants.

Factors Influencing Detergency

Detergency is mainly affected by the concentration and structure of surfactant, hardness and builders present, and the nature of the soil and substrate. Other important factors include wash temperature; length of time of washing process; mechanical action; relative amounts of soil, substrate, and bath, generally expressed as the bath ratio, ie, the ratio of the bath weight to substrate weight; and rinse conditions.

Effect of Surfactant Concentration. A plot of soil removal versus surfactant concentration is generally sigmoid. It starts at the soil removal of water without surfactant, rises slowly, then more steeply until a plateau is reached when detergency is little affected by increase in surfactant concentration. This plateau can be correlated with the critical micelle concentration, CMC, of the surfactant, and is generally higher than the CMC. In general, detergency attains its maximum when the CMC of the surfactant is reached, taking into account surfactant adsorption on the soil and substrate. However, with certain surfactants, there is evidence that oily-soil detergency continues to increase above the CMC (17–20).

Surfactant Structure. The chemical structure of the surfactant is an important factor in deterative effectiveness. When relating deterative power to chemical constitution, within limited series certain regularities can be observed, but few if any general principles apply to the whole range of surfactants. Among the homologous fatty acid soaps and the straight-chain alkyl sulfates, optimum detergency under usual washing conditions occurs at a chain length of ca sixteen carbon atoms. Detergency among the ethoxylated nonionic surfactants varies in a regular manner with the length of the ethylene oxide chain as well as with the structure of the hydrophobic group. In general, optimum detergency occurs with 12–16 carbon atoms in the hydrophobe chain, and a hydrophile–lipophile balance, HLB, of about 12. For oily soil detergency by nonionics it is often found the optimum corresponds with the phase inversion temperature of the system under consideration (21).

Within a series with a fixed hydrophilic head group, detergency increases with increasing carbon chain length, reaches a maximum, and then decreases. This behavior frequently reflects a balance between increased surface activity of the monomer and decreased monomer concentration with increased surface activity. Similar effects are seen in surfactants in biological systems.

The numerous studies of the effect of surfactant structure in detergency include the classical paper on the series of sulfated secondary straight-chain alcohols (22) and various papers on detergent–builder combinations (23); nonionic–anionic mixtures (24,25); the polyethenoxy nonionic series (21,26,27); detergency of isomeric alkylbenzenesulfonates (28); practical home laundering (29); laundering and detergency effects in seawater (30); *o*- and *p*-alkylbenzene-sulfonates with straight and branched chains (31); and studies on alkylbenzene-sulfonates (32).

Water Hardness and Builders. The presence of heavy-metal ions, especially calcium and magnesium, has an effect on washing second only to that of the surfactant itself. Distilled water, used in a system where soil and substrate do not contain substantial calcium ions, is a surprisingly effective detergent, when the soils themselves contain fatty acids or sodium soaps. Detergency can be improved by the addition of surfactants in amounts smaller than might be expected. Conversely, the detergency of surfactants is generally decreased by the presence of hardness ions (Ca²⁺ and Mg²⁺) and hard water alone is a poor soil remover. With traditional formulations containing anionic surfactants or soaps, cotton can be cleaned well by washing only if the calcium concentration is reduced to <0.01 mM. In current synthetic detergent compositions, builders sequester calcium and magnesium ions and thus reduce interaction with surfactants, soils, and fabrics as described above.

The sodium soaps of fatty acid form calcium soaps of such low solubility that they act as their own builders. Initial soap additions precipitate the calcium ion and the soap added thereafter functions in soft water. At high temperatures, the calcium soaps are relatively soluble compared to calcium tripolyphosphate. Thus sodium tripolyphosphate (STPP) can build (revert) soaps in a hot water wash. However, at low temperatures the relative affinity of STPP for calcium decreases so that STPP cannot build soaps in a cold water wash.

Calcium ion enters the system not only in the form of water hardness but also in the form of calcium salts contained in the soil. Other heavy-metal ions such as aluminum and ferric iron may also be present in the soil, and must be removed by an appropriate builder to achieve good soil removal. Effective builders for cotton washing are those for which the calcium dissociation constant, expressed as p*K*_{Ca}, or –log *K*_{Ca} is >4 and preferably >7 (33). Much of the work that led to elucidating the role of builders as calcium sequestering agents in detergency was done in connection with redeposition studies (34).

Legislatively mandated reductions in detergent phosphate concentrations have resulted in numerous attempts to compensate for the attendant cleaning losses. Problems caused by phosphate reduction can be ameliorated by changes in surfactant systems. Thus, calcium-sensitive surfactants such as linear alkyl-benzenesulfonates (LAS) can be replaced by calcium insensitive ones such as alcohol ethoxylates (AE) or alcohol ether sulfates (AES). Proper blends greatly reduce surfactant sensitivity to calcium ions (35,36). Increasing the amount of surfactants partially compensates for phosphate reduction or elimination. Other calcium-lowering agents include NTA, citrate, zeolite, and carbonate. Citrate and NTA lower calcium levels in solution by sequestration, zeolite by ion exchange, and carbonate by ion pair formation and precipitation. To date, no entirely satisfactory single replacement for phosphate has been found. Indeed, some soil-cloth detergency data show that the 1978 nonphosphate products are less effective than the 1969 high-phosphate products (37). However, subsequent refinements in formulation technology have substantially improved nonphosphate product performance to equal that of their phosphate predecessors.

In the presynthetic surfactant era, soap was built (and still is) with alkaline salts such as soda ash, silicates, orthophosphates, and borates. These materials buffer the wash solution to a high pH and prevent soap protonation; thus the soap remains effective. Another type of builder is the neutral

inorganic salt such as sodium chloride and sodium sulfate. These materials may improve detergency by increasing the ionic strength and altering the CMC of anionic surfactants.

Antiredeposition Agents. The redeposition effect is best illustrated by agitating a clean swatch of white cotton in a washing bath that has been used and is turbid with dispersed soil. Even though the bath may still be effective and capable of removing soil, the white swatch picks up some soil from the bath and darkens. Any agent that minimizes this redeposition improves net soil removal, and decreases the soil accumulation through several laundering cycles. Redeposition is in many ways the inverse of soil removal and is influenced by the same factors. Thus, calcium ion promotes redeposition in much the same way that it inhibits soil removal. Surfactants themselves are effective antiredeposition agents. Builders reduce redeposition independently of their ability to sequester hardness ions.

Sodium carboxymethylcellulose, NaCMC, greatly reduces redeposition in cotton-washing systems based on synthetic surfactants. It is effective at remarkably low concentrations of ca 1% of the standard washing compositions used at ca 0.1 to 0.2% in the bath. Thus, ca 0.001–0.002%, or 10–20 ppm NaCMC is sufficient to significantly inhibit redeposition.

A very large number of hydrophilic polymers have been tested with regard to their power of inhibiting redeposition, and it is remarkable how few materials can match NaCMC. Some proteins, polyvinylpyrrolidinones, and vinyl alcohol polymers with rather specific molecular weight ranges are in the same class as NaCMC. Certain starch derivatives, cellulose sulfates, and cellulose ethers are also effective. Currently, both NaCMC and cellulose ethers are used in home laundry products specifically to control redeposition. The cellulose ethers are used to control oily soil redeposition upon synthetic fibers, especially polyester. Using tracer methods, it was proved that NaCMC is adsorbed on the soil particles and also, under normal laundering conditions, on cotton fiber (38,39). Adsorption on cotton appears to be more important in preventing soil redeposition. This absorption occurs to a significant extent only in the presence of metallic cations that are furnished as sodium ions from STPP and the filler salts present in the detergent. In the absence of cations, except for the Na⁺ present in the NaCMC itself, the adsorption is so slight as to be almost undetectable, and the antiredeposition effect is correspondingly lowered.

Liquid Soil. Liquid soils are frequently removed from fibers by a roll-up mechanism (40). In a study of wool grease removed from single fibers, it was found that the grease layer, originally continuous, rolled up in relatively large globules that detached themselves from the fiber. The detergent alters the contact angle at the fiber–grease–water interface. In pure water, the contact angle is <90°, measured through the oil. Addition of detergent increases the angle to ca 180°. The receding contact angle, ie, that measured as the oil rolls up, is more important in detergency than is the advancing contact angle. Detergency mechanisms are similar on wool, viscose rayon, and cuprammonium rayon (41). The soil removal depends primarily on the nature of the oil. Those containing free fatty acids are removed more rapidly, followed by neutral glycerides and less polar mineral oils. Similarly, polar oils are more easily removed from wool than nonpolar oils (42). This difference correlates well with differences in contact angle.

Liquid soil can be removed by direct emulsification and solubilization as well as by roll-up. Oils that are more readily emulsified or solubilized are removed more rapidly and more completely.

Solid Soil Type and Size. Different solid soils differ greatly in ease of removal and redeposition behavior. These differences can be traced to particle size and soil–substrate bonding. The effect of particle size variation on detergency has been studied with soil removal and redeposition techniques.

In an early systematic investigation of this relationship, a series of carbon black samples was applied to chopped cotton fibers from an aqueous suspension with vigorous agitation (43). The degree of soiling was estimated by filtering the fibers with suction to form a mat, drying, and measuring the reflectance of the mat. The degree of soiling was found to be directly related to the particle size of the carbon black. As shown in Figure 1, when the primary particles of the carbon black are <50 nm, the soil deposition is severe, whereas at particle sizes >50 nm, the soil deposition is much lighter and relatively independent of particle size. Electron microscope studies of the cotton fiber show that 50 nm is about the upper limit of the crevice width on the surface of cotton fibers. Furthermore, the soil deposition by this procedure using fine-particle carbon blacks is practically irremovable, and the fibers are to all intents and purposes dyed a very fast gray. On this basis and on the basis of other supporting evidence, it was concluded that soiling in this system takes place by microocclusion, ie, that the particles of soil are so small that they can be trapped mechanically by the irregularities of the fiber surface, even though these irregularities are scarcely larger than the dimension of a polymeric molecule. This theory accounts well for the observed facts, but the facts could also be accounted for by simply assuming that van der Waals forces control the attachment of the carbon particles to the fiber. Since these forces operate at the surface of the particle, there should be a critical particle size below which the surface forces become more important than the inertial and gravitational forces that are proportional to mass. Electron micrographs of soiled cotton show adherent soil particles in size ranges up to 0.5–2.0 μm clinging to apparently smooth surfaces (44–46). These and other supporting detergency data indicate that mechanical binding is not the basic mechanism of solid soil binding, even though the binding varies greatly with particle size.

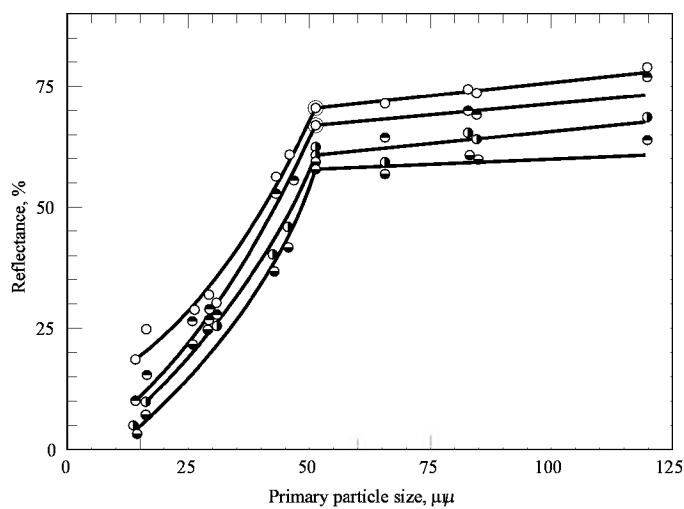


Fig. 1. Effect of primary particle size in a 1% carbon black dispersion on rate of change in reflectance. ○ 5 min; ◐ 20 min; ● 180 min; ● 360 min.

As might be expected, large differences in the removability of solid particulate soil are due to differences in the chemical nature of the particle surface. Thus, iron oxides, lampblacks, and clays, all of the same particle size, differ greatly in their redeposition behavior and the manner in which they are removed.

Fabrics and Fibrous Substrates. Certain fibers are easier to wash than others. The hardness of the fiber surface, which varies not only with the basic fiber but also with the surface finish, affects both soilability and soil removal. Generally, soft finishes pick up and retain soil more readily than hard finishes, since soft surfaces permit greater soil–surface contact. The hydrophobicity of the fiber surface also influences soilability and soil removal, and the more hydrophobic fibers show greater soil retention especially for hydrophobic soils such as oils. These effects are explained by the high-energy interface existing between hydrophobic fibers and water (47). Many soil components lower the interfacial energy and therefore locate at the fiber–water interface.

Soils can also penetrate into the fiber. The interior of the cotton (qv) fiber, the lumen, is relatively hollow, and soils may collect there. Polyester fibers

are solid, but if polyester is washed above its glass-transition temperature T_g it becomes relatively fluid (see FIBERS, POLYESTER). In this case, oils on the fiber surface can mix with the polyester itself. When soils penetrate into the fiber, they are nearly impossible to remove.

Temperature and Mechanical Action. Absorption of external thermal or mechanical energy by a deterative system influences the rate and extent of soil removal. Raising the temperature generally increases the cleaning rate and, therefore, the amount of soil removed during any fixed-time laundering cycle. The effect is only strong at two critical temperatures; one is the temperature where the fatty soil in the system liquefies. As temperature increases through this region, the detergency increases markedly. Typical curves of detergency versus temperature are shown in Figure 2 (48). The second critical temperature is the boiling point. Boiling greatly increases detergency because of the localized mechanical action of steam bubbles forming, expanding, and breaking away at the solid-liquid interfaces. However, in the United States, boiling effects are never encountered in home washing machines. U.S. laundry wash temperatures have declined to an average wash temperature of about 35°C, and the average hot water wash temp is only 54°C. From 1970 to 1988 hot water wash declined from 48 to 20% of washes (49). Temperature effects are also important with certain detergent additives. Thus the efficacy of certain bleaches and enzymes is markedly reduced at the low wash temperatures now popular. These effects emphasize the problems inherent in the current trend to lower wash temperatures.

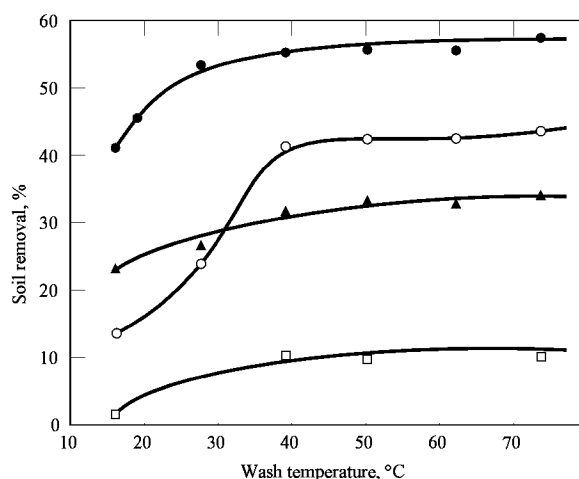


Fig. 2. Effect of wash temperature on removal of ● sebum, ○ lanolin, and ▲ lard from cotton (0.25% built detergent) (48). □, Represents sebum removal with no detergent.

As with the case of energy input, detergency generally reaches a plateau after a certain wash time as would be expected from a kinetic analysis. In a practical system, each of its numerous components has a different rate constant, hence its rate behavior generally does not exhibit any simple pattern. Many attempts have been made to fit soil removal (50) rates in practical systems to the usual rate equations of physical chemistry. The rate of soil removal in the Launder-Ometer could be reasonably well described by the equation of a first-order chemical reaction, ie, the rate was proportional to the amount of removable soil remaining on the fabric (51,52). In a study of soil removal rates from artificially soiled fabrics in the Terg-O-Tometer, the percent soil removal increased linearly with the log of cumulative wash time.

In detailed studies with a number of different artificial test cloths, first-order kinetics were obtained for soil removal during the first 6–20 min of the cycle (53).

In the washing of fabrics, machine and timing must be carefully adjusted. Too large an energy input for too long a time can injure fabric severely, and eventually tear it to shreds. The same effect is quite possible with purely mechanical agitation.

Foam. The relationship of foaming power to detergency has always been of interest, and foaming power has become associated in many consumers' minds with high deterative power (see also FOAMS). However, foam has no direct relationship to detergency in ordinary fabric-washing systems, and does not improve cleaning in a laundry or home-washing machine. Indeed, excessive foam can inhibit agitation and reduce cleaning. Additionally, excess foam levels may concentrate certain surface active cleaning agents in the foam, reducing their contact with the fabrics to be cleaned. In systems of low bath ratio, on the other hand, foam may play an important role. The individual foam laminae tend to imbibe and segregate particles of both liquid and solid soil that have been removed from the substrate. This prevents redeposition and enables the soil to be easily removed by scraping off or rinsing away the soil-laden foam. This effect is very important in the on-location shampooing of rugs, and to a certain extent in the shampooing of human hair. An excellent series of photomicrographs show the imbibition of liquid-soil droplets by foam films during the washing of oiled glass plates (54). The soil tends to be carried by capillary convention to the low-pressure region in plateau borders where three individual laminae join to form the edge of a foam cell.

Mechanisms

Even the simplest deterative system is surprisingly complex and heterogeneous. It can nevertheless be conceptually resolved into simpler systems that are amenable to theoretical treatment and understanding. These simpler systems are represented by models for substrate-solid soil and substrate-liquid soil. In practice, many soil systems include solid-liquid mixtures. However, removal of these systems can generally be analyzed in terms of the two simpler model systems. Although these two systems differ markedly in behavior and structure, and require separate treatment, there are certain overriding principles that apply to both.

The first principle is that soil systems can be regarded and treated as classical systems of colloid and surface chemistry. A cotton fiber bearing attached small particles of clay or carbon soil, immersed in an aqueous medium, is fully analogous to the sulfur sols or gold sols of classical colloid chemistry (see COLLOIDS). Although the fiber is not of colloidal dimensions, the area over which any individual soil particle contacts the fiber is of colloidal dimensions. Furthermore, the ratio of the soil particle's area to the area of contact, and the ratio of the soil particle's mass to the area of contact put this system in the colloidal range. These ratios are the same as the ratios prevailing when one sulfur sol particle makes contact with another. In addition, in the substrate-soil system two different types of surfaces are interacting, whereas in the classical sol systems the particles are all of the same chemical nature and the interacting surfaces are therefore similar. The problem of mixed sols has, however, been explored in classical colloid chemistry to at least a limited extent, and it is fully analogous to the problem of the model soil-deterative system. A fiber or steel plate covered with oil and immersed in an aqueous solution is a good representative of the liquid-liquid-solid system of classical surface chemistry. The well-established concepts of interaction at the liquid-liquid interface and at the three-phase boundary line are fully applicable.

A second principle applying to these model systems is derived from their colloidal nature. With the usual thermodynamic parameters fixed, the systems come to a steady state in which they are either agglomerated or dispersed. No dynamic equilibrium exists between dispersed and agglomerated states. In the solid-soil systems, the particles (provided they are monodisperse, ie, all of the same size and shape) either adhere to the substrate or separate from it. In the liquid-soil systems, the soil assumes a definite contact angle with the substrate, which may be anywhere from 0° (complete coverage of the substrate) to 180° (complete detachment). The governing thermodynamic parameters include pressure, temperature, concentration of dissolved components, and electrical conditions. This concept is at variance with the idea of a dynamic detergency equilibrium which was advanced as a working hypothesis by many of the early investigators. It is however, much more in keeping with the established concepts of colloid chemistry. In the classical theory of lyophobic colloids, no consideration is given to a dynamic equilibrium existing between the peptized state and the agglomerated state. This

all-or-nothing behavior has been studied experimentally in model deterive systems by observing, with the help of microscopic techniques, the behavior of single textile fibers in aqueous carbon black suspensions containing dissolved soap (55). In the absence of soap, the carbon particles deposit on the fiber. When sufficient soap is added, the carbon particles do not deposit on the fiber but remain in suspension. The quantity of soap necessary to maintain the carbon in suspension is proportional to the amount of carbon present in the system. There is a sharp demarcation between the state in which the carbon is suspended and the state in which the carbon is attached to the fiber.

In applying this concept, the factor of particle size must be continuously borne in mind. A heterodisperse system can reach a steady state wherein the smaller particles are agglomerated and the larger particles are dispersed, giving the apparent effect of an equilibrium. In ideal monodisperse systems under steady conditions, however, no such effects are noted.

Purely mechanical disturbances (which are not usually considered thermodynamic variables) may influence the state of aggregation of a colloidal system; for example, floc size in carbon and iron oxide suspensions varies with the degree of agitation being imposed on the system (56). When the agitation is stopped, the flocs revert to their steady-state size. To have any effect, the shear field must be quite high and the particles relatively large. With particles <0.5 μm, thermal agitation becomes more important than even vigorous mechanical agitation in determining the state of aggregation.

A final consideration in resolving practical deterive systems into their simpler components relates to soil removal versus redeposition. Superficially, it would appear that the redeposition phenomenon contradicts the all-or-nothing concept that the system must exist in either the agglomerated or the dispersed state. Keeping in mind both the composite nature and the kinetics of a practical system, it is readily shown that no such contradiction exists. The soil particle that redeposits is in a different state from what it was during its initial removal from the substrate. It may have changed its state by becoming detached from adherent oil or from a cluster of similar solid particles. During the time interval between detachment and redeposition, it may have altered its surface character by adsorption or desorption. In this time interval, the state of the substrate surface may also have changed by adsorption or desorption, or by the loss or gain of an oily layer. Additionally, there are probably varying substrate types present in the wash, with varying affinities for the soils present. For example, oily soil could be detached from a cotton fiber, but then deposit on a polyester fiber for which it would have a greater affinity. Thus, the initial group of agglomerated systems, which composes the soil–substrate–bath complex before soil removal, is quite different from the agglomerated system composed of substrate and redeposited soil.

SOLID-SOIL DETERGENCY

Adsorption. Many studies have been made of the adsorption of soaps and synthetic surfactants on fibers in an attempt to relate detergency behavior to adsorption effects. Relatively fewer studies have been made of the adsorption of surfactants by soils (57). Plots of the adsorption of sodium soaps by a series of carbon blacks and charcoals show that the fatty acid and the alkali are adsorbed independently, within limits, although the presence of excess alkali reduces the sorption of total fatty acids (58). No straightforward relationship was noted between detergency and adsorption.

In a study of the adsorption of soap and several synthetic surfactants on a variety of textile fibers, it was found that cotton and nylon adsorbed less surfactant than wool under comparable conditions (59). Among the various surfactants, the cationic types were adsorbed to the greatest extent, whereas nonionic types were adsorbed least. The adsorption of nonionic surfactants decreased with increasing length of the polyoxyethylene chain. When soaps were adsorbed, the fatty acid and the alkali behaved more or less independently just as they did when adsorbed on carbon. The adsorption of sodium oleate by cotton has been shown independently to result in the deposition of acid soap (a composition intermediate between the free fatty acid and the sodium salt), if no heavy-metal ions are present in the system (60). In hard water, the adsorbate has large proportions of lime soap.

The adsorption behavior of commercial alkylbenzenesulfonates and several other anionic detergents on cotton, wool, and some of the synthetic fibers has been carefully plotted (61). The equilibrium adsorption isotherms, with a few exceptions, do not show any startling abnormalities. A number of other studies have attempted to correlate adsorption with detergency (62–64). In general, no significant correlations were found for a number of possible reasons. On porous fibers such as cotton, adsorption may occur on fiber areas that are not beneficial to detergency. The presence of soils on the fiber surface may significantly alter the adsorption behavior. The presence of trace amounts of surface-active agents in the soils themselves could modify adsorption of the active detergent component. For instance, soap present in natural soils could significantly change nonionic adsorption behavior. Finally, and perhaps most importantly, equilibrium adsorption measurements may not realistically reflect adsorption taking place in a 10 min wash cycle.

Adsorption of bath components is a necessary and possibly the most important and fundamental detergency effect. Adsorption (qv) is the mechanism whereby the interfacial free energy values between the bath and the solid components (solid soil and substrate) of the system are lowered, thereby increasing the tendency of the bath to separate the solid components from one another. Furthermore, the solid components acquire electrical charges that tend to keep them separated, or acquire a layer of strongly solvated radicals that have the same effect. If it were possible to follow the adsorption effects in a deterive system, in all their complex ramifications and interactions, the molecular picture of soil removal would be greatly clarified.

Mass Transfer near the Substrate Surface. Mechanical action has a great effect on soil removal, probably by influencing mass transfer (qv), ie, the diffusion of soluble material away from the immersed fibers (65,66). Mechanical action tends to maintain a high concentration gradient near the fiber, and the resulting increased diffusion causes stronger diffusion currents to flow. These diffusion currents are presumably responsible for carrying away the soil particles that have already been detached or loosened from the fiber surface by physicochemical action. In a Terg-O-Tometer investigation of mass transfer, using only water soluble substances, the transfer coefficient was found to be directly proportional to agitator speed and stroke angle, inversely proportional to the water-holding capacity of the cloth load, and independent of bath volume (65,66). The soil removal behavior is very similar to the mass transfer behavior, supporting the idea that diffusion currents are an important operating factor in soil removal, even though they may play no part in breaking the primary soil–substrate bond. Considering the wide variety of soils and substrates encountered in washing, as well as the wide variation in particle size and fabric geometry, mass transfer effects could play a principal or minor role of the overall wash process, depending upon the specific system encountered.

Colloidal Stabilization. Surfactant adsorption reduces soil–substrate interactions and facilitates soil removal. For a better understanding of these interactions, a consideration of colloidal forces is required.

The model solid-soil deterive system is advantageously treated as a sol-agglomerate colloid system or, in more general terms, a lyophobic colloid. In the typical lyophobic colloid, consisting of a single disperse phase in an aqueous suspending medium, only one type of liquid-solid interface and one type of solid–solid interface is present. The simplest deterive system, however, has an added degree of complexity in the presence of two types of liquid–solid interface: soil–bath and substrate–bath. Also present are two effective types of solid–solid interface: soil–substrate and soil–soil. The soil–soil interface relates to flocculation or dispersion of soil particles remote from the substrate, and is not of primary concern in the present discussion. The soil–soil interface is, of course, important in practical detergency since soil aggregates can be regarded as large single particles.

There are two general theories of the stability of lyophobic colloids, or, more precisely, two general mechanisms controlling the dispersion and flocculation of these colloids. Both theories regard adsorption of dissolved species as a key process in stabilization. However, one theory is based on a consideration of ionic forces near the interface, whereas the other is based on steric forces. The two theories complement each other and are in no sense contradictory. In some systems, one mechanism may be predominant, and in others both mechanisms may operate simultaneously. The fundamental kinetic considerations common to both theories are based on Smoluchowski's classical theory of the coagulation of colloids.

As the particles in a colloidal dispersion diffuse, they collide with one another. In the simplest case, every collision between two particles results in the formation of one agglomerated particle, ie, there is no energy barrier to agglomeration. Applying Smoluchowski's theory to this system, the half-life, $t_{1/2}$, ie, the time for the number of particles to become halved, is expressed as follows, where η is the viscosity of the medium, k Boltzmann's constant; T temperature; and N_0 is the initial number of particles.

$$t_{1/2} = \frac{3\eta}{4kTN_0}$$

For dispersions at moderate concentrate, ca 10⁹ particles per cm³, $t_{1/2}$ has a value of the order of a few seconds. This expression assumes there is no barrier to collision and every collision is effective. For stable dispersions to exist, an energy barrier W is assumed that prevents collision. In this case, the expression for half-life becomes:

$$t_{1/2} = \frac{3\eta}{4kTN_0} \exp(2W/kT)$$

If W is 15–20 kT , the half-life is several days and the dispersion is reasonably stable. The problem of colloidal stability thus becomes a problem of the energy barrier, W , that lessens the unifying collisions.

Ionic Stabilization. The quantitative theory of colloid stability based on electrical barriers was developed independently in the USSR by Derjaguin and Landau (67) and by Verwey and Overbeek in Holland (68). It is generally referred to as DLVO theory. There are two opposing forces between particles: a force of attraction that is responsible for agglomeration when the particles approach close enough, and a force of repulsion that prevents close approach. The attractive forces are the London dispersion forces or van der Waals forces. These have the general form shown below, where r is the distance between particles and n is between 3 and 7, depending on the geometry of the interacting surfaces.

$$F = K/r^n$$

The repelling forces are caused by the electrical double layers that surround the particles in aqueous dispersions. The interaction of these two forces is conveniently shown in a plot of potential energy versus distance between particle centers, as in Figure 3. The ordinate in this plot can be either force or potential; for quantitative purposes, they are interconvertible through the definitive relationship that force is the negative gradient of potential. To separate two agglomerated particles ($r = 2a$), it is necessary to overcome the energy barrier $V + W$, to bring them to distance X_2 at which point repulsive forces predominate. Conversely, to cause agglomeration of two separated particles, it is necessary to overcome the barrier W at X_2 .

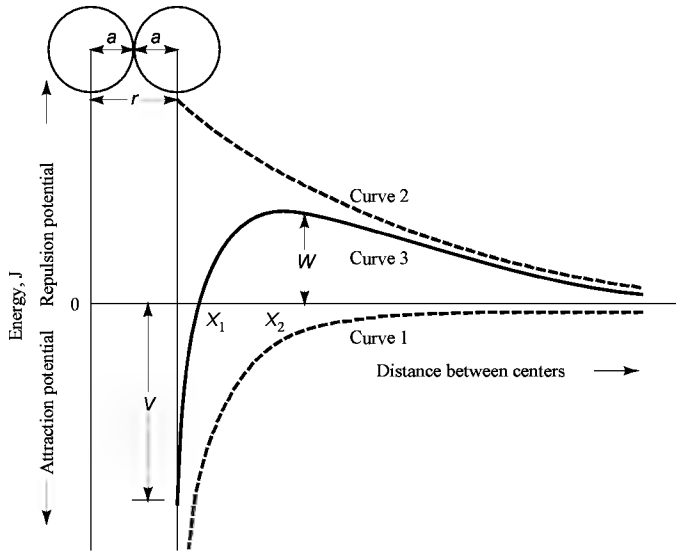


Fig. 3. Attraction–repulsion potentials as a function of distance between particle centers. Curve 1 represents the attractive potential caused by van der Waals forces, curve 2 is the repulsive potential caused by double-layer forces, and curve 3 is the resultant force experienced by the two particles.

According to DLVO theory, the shape of curve 2, which controls the shape of the effective curve 3, is determined by the thickness and charge density of the electrical double layer that surrounds each particle. These factors, in turn, are determined by the nature of the ionic species adsorbed at the interface, the degree of adsorption, and the ionic strength of the surrounding medium. The quantitative aspects of the DLVO and the double-layer theory (ie, the structure, formation, and interactions of electrical double layers at phase interfaces) are beyond the scope of this article, and it is recommended that the reader consult the standard publications in this field. Qualitatively, high charge density increases the repulsion forces and thus shifts curve 3 upward, increasing the potential W and decreasing $|V|$. High ionic strength shrinks the double layer, lowering curve 3, lowering W , and tending to increase $|V|$.

Electrical potentials can, of course, be negative or positive. In most detergent systems, the charges acquired by both substrate and solid soil are negative. The adsorption of detergent anions increases the effective density of negative charge and tends to increase the double-layer-repulsion potentials. The counterions, in this case metallic cations, exert the controlling effect on the ionic strength factor, with increasing valence greatly magnifying the effect. Thus, calcium ions have a much greater effect than sodium ions in shrinking the double layer and promoting agglomeration. Aluminum ions have an even greater effect than calcium ions. This is in accord with the well-known Schulze–Hardy rule that was proposed as an empirical finding many years before this quantitative DLVO theory was developed.

The electrokinetic effect is one of the few experimental methods for estimating double-layer potentials. If two electrodes are placed in a colloidal suspension, and a voltage is impressed across them, the particles move toward the electrode of opposite charge. For nonconducting solid spherical particles, the equation controlling this motion is presented below, where u = velocity of particles; η = viscosity of medium; V = applied field, V cm; R = factor for electrical relaxation; D = dielectric constant of medium; F = factor for size of spheres; and ζ = zeta potential.

$$u = \frac{RFVD}{6\pi\eta}\zeta$$

This equation is a reasonable model of electrokinetic behavior, although for theoretical studies many possible corrections must be considered. Correction must always be made for electrokinetic effects at the wall of the cell, since this wall also carries a double layer. There are corrections for the motion of solvated ions through the medium, surface and bulk conductivity of the particles, nonspherical shape of the particles, etc. The parameter zeta, determined by measuring the particle velocity and substituting in the above equation, is a measure of the potential at the so-called surface of shear, ie, the surface dividing the moving particle and its adherent layer of solution from the stationary bulk of the solution. This surface of shear lies at an indeterminate distance from the true particle surface. Thus, the measured zeta potential can be related only semiquantitatively to the curves of Figure 3.

There have been many attempts to correlate detergency and deflocculation with measurements of zeta potential (69); such measurements have been made on both fibers and soil particles using various model detergent baths as the media. Aside from the expected result that high zeta potentials tend to correlate with deflocculation, few direct relationships of theoretical significance have become evident. In an extensive and productive study, it was pointed out that DLVO theory relates the height of the maximum in the potential curve (W in curve 3, Fig. 3) to the square of the zeta potential rather than to its first power (70). From the DLVO theory the surface potential ψ was calculated, which is necessary to give a barrier of height (in joules) $W = 15 kT$ for soil particles of various size in electrolytes of various concentrations; ζ was measured experimentally. When the ratio ζ^2/ψ^2 was greater than unity, good stability was attained. In calculating soil–substrate interaction, good results were obtained using the ratio of $[(\zeta_{\text{substrate}})(\zeta_{\text{soil}})]/[\psi_{\text{substrate}})(\psi_{\text{soil}})]$. Application of DLVO theory to detergency has also been made very successfully when the shape factors and factors involving the structure of the double

layer were taken into consideration (71).

The strong adverse influence of calcium ions on the stability of lyophobic suspensions is predicted by DLVO theory, and has been demonstrated with many types of simple soils. That calcium ions have an overwhelming effect on the redeposition of carbon soil onto cotton tends to support the idea that DLVO theory is a principal key in explaining deterative action. The redeposition of carbon onto cotton has been correlated quantitatively with the calcium ion content of the system, both in the presence and absence of surfactant (72). The adverse effect of calcium ions on wet soil removal in practical washing has also been well established (73). The effect of calcium in detergency cannot be explained solely, however, by its shrinking of the double layer. Calcium is very strongly adsorbed onto cotton through the carboxyl and hydroxyl groups present on the cellulose chains. This adsorption is so strong that cotton rinsed in even slightly hard water transfers a harmful amount of calcium to the next wash liquor. Calcium present in soil particles can act as a bonding agent at the soil–substrate interface, and inhibit soil removal by this secondary mechanism. In one of the most thorough studies of the above effects, it was also demonstrated that calcium promotes the deposition of oil onto cotton (74). The sorption of calcium during washing is complicated by its apparent ability to coadsorb with surfactants and complex phosphate builders, further obscuring the mechanism by which it exerts its adverse effect (75,76).

Despite its successes and its great usefulness as a guide, the limitations of the DLVO theory in explaining detergency are most evident when considering the nature of the soil–substrate bond. The DLVO theory in its simplest form postulates van der Waals' bonding as the primary agglomerating mechanism. Even revised theories of double layer interaction, that do not assign such an important role to van der Waals' forces, do not encompass agglomerating mechanisms such as polyvalent cation bridges and hydrogen bonding (77,78). There is ample evidence that these latter mechanisms prevail in some important deterative systems.

Steric Stabilization. Double-layer repulsion cannot satisfactorily explain stabilization of lyophobic suspensions by nonionic surfactants nor nonionic detergency in solutions of high salt content. Not only nonionic surfactants but even anionic ones can overcome the agglomerating effects of salt. To explain these phenomena, steric interaction between the surfactant molecules adsorbed at the solid–solution interface must be considered. The adsorbed molecules are oriented with the hydrocarbon tails adherent to the essentially hydrophobic solid surface and the water-soluble heads sticking into the solution. This adsorption and orientation is governed by the same factors governing micelle formation, ie, the amphipathic nature of the surfactant molecule and solution forces. The head groups at both nonionic and anionic surfactants are heavily hydrated. As a result the lyophilic particle is surrounded by an inner layer of hydrophobic surfactant tails and an outer layer of hydrated surfactant heads. In Figure 3, it can be seen that the attractive potential does not greatly increase until the particles are relatively close to each other.

The bulk of the hydrated head groups prevents the particles from approaching near enough to each other for the attractive forces to cause agglomeration. In the case of polyoxyethylene nonionic surfactants, the head group is quite large, extending for several nanometers into solution; the head group is associated with a large amount of water of hydration caused by ether–water interaction. The bulk of a hydrated anionic head group is not so great; however, the anionic surfactant retains some ionic stabilization even in salt solution. Agglomeration is achieved only if the particles overcome the energies involved in steric compression of the head groups and desolvation of the head groups. Conversely, if two particles, adherent over a small area of contact, are immersed in a surfactant solution, oriented adsorption on the noncontacting areas, followed by hydration, presumably breaks the adhesive bond and forces the particles apart. This picture would suggest some degree of correlation between the wetting power of a surfactant for a given material and its stabilizing power for sols of the same material. Such a correlation has indeed been found, using paraffin wax with both nonionic and anionic surfactants (79).

OILY-SOIL DETERGENCY

Roll-up. The principal means by which oily soil is removed is probably roll-up. The applicable theory is simply the theory of wetting. In briefest outline, a droplet of oily soil attached to the substrate forms at equilibrium a definite contact angle at the oil–solid–air boundary line. This contact angle (Fig. 4) is the result of the interaction of interfacial forces in the three phase boundaries of the system. These interfacial forces, expressed in mN m(dyn cm), or interfacial free energy values expressed in mJ m² (erg cm²s) are conveniently designated γ_{LA} , γ_{SA} , and γ_{SL} ; the subscripts relate to the liquid–air, solid–air, and solid–liquid interfaces, respectively. The equilibrium contact angle at the boundary line θ_{SLA} may be regarded as a thermodynamic quantity since it is functionally related to the free energy values through the Young-Dupre equation:

$$\gamma_{SA} = \gamma_{SL} + \gamma_{LA} \cos \theta_{SLA}$$

When the solid substrate is placed in the bath, the air is displaced by the bath, *B*, and the *SA* interface is replaced by an *SB* interface. Similarly, an *LB* interface replaces the *LA* interface. The equilibrium free energy values of these new interfaces are not established immediately but gradually through mass transfer (if there is any mutual solubility between *L* and *B*; it is assumed that *B* does not dissolve *S*) and through adsorption of dissolved components. When these processes have gone to completion the new relationship is

$$\gamma_{SB} = \gamma_{SL} + \gamma_{LB} \cos \theta_{SLB}$$

In general, γ_{SB} and γ_{LB} are lower than γ_{SA} and γ_{LA} . If *B* contains a surface-active agent, they tend to become exceptionally low and θ_{SLB} is therefore much greater than θ_{SLA} . Because θ_{SLB} is greater than 90°, even though less than 180°, hydraulic action is capable of removing the oil droplet from the soil and surface.

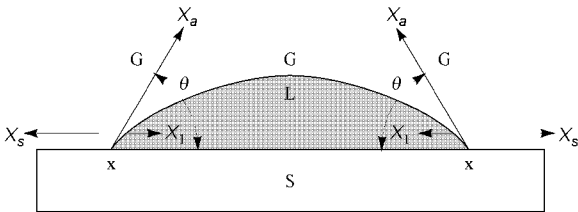


Fig. 4. Contact angle at the oil–solid–air boundary line in the roll-back process of oily solid detergency. X_s , X_l , and X_a are force vectors at the surface and tangential to the droplet.

Rearrangement of the above equation gives

$$\cos \theta = \frac{\gamma_{SB} - \gamma_{SL}}{\gamma_{LB}}$$

For $\theta > 90^\circ$, there must be $\cos \theta < 0$. Since the interfacial tension terms are themselves positive, for $\cos \theta < 0$, it follows that $\gamma_{SL} > \gamma_{SB}$. In other words, for effective roll-up, ie, $\theta > 90^\circ$, the interfacial tension between the solid and bath must be less than that between the solid and liquid. As shown in Figure 5, the area of contact can be reduced to zero while maintaining the contact angle and its equilibrium value, and avoiding any necking down and division of the droplet. If θ_{SLB} is less than 90°, it is impossible to separate the oil completely from the surface by hydraulic action alone. As the droplet is withdrawn and the area of contact is reduced, a neck is formed directly above the contact region. If the contact angle is maintained at equilibrium, the drop must divide at this neck, leaving a small quantity of oil adherent to the substrate. This effect is shown in Figure 6.

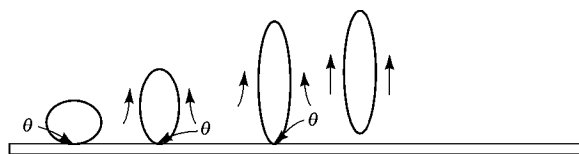


Fig. 5. With θ remaining constant at $>90^\circ$, oil droplet can be removed completely by hydraulic currents (arrows).

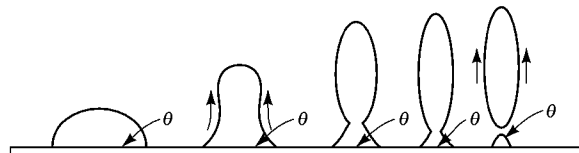


Fig. 6. With θ remaining constant at $<90^\circ$, droplet cannot be removed completely by hydraulic currents (arrows). A small droplet is left attached to the substrate.

The value of θ_{SLB} can be estimated on purely theoretical grounds from estimates of the adsorption of surfactant which, in turn, can be estimated from the Gibbs adsorption equation relating adsorption to surface-tension lowering.

Even when the equilibrium value of θ_{SLB} approaches 180° , it is quite difficult in practice to displace all the oil from the substrate by the bath for the following reasons: the contact angle of the oil, as it is rolling back, is a receding contact angle that is considerably smaller than the equilibrium angle. Any surface roughness further increases the hysteresis. Thus, an actual angle of 180° may not be achieved even though the theoretical equilibrium angle has this maximum value. Solid surfaces are notoriously inhomogeneous. Even a small surface area exhibits roughness or spots of high interfacial energy where the moving boundary tends to stick. The hydraulic currents then tear away the bulk of the droplet, leaving some oil at the rough or sticky spot.

In this model, the surface is oil-free under steady-state conditions only when θ_{SLB} is 180° . At any lower value, free-floating oil droplets redeposit and become attached to the substrate, and at equilibrium (on the all-or-nothing basis) all the oil is attached. When θ_{SLB} is high, however, even though it is not 180° , agitation keeps a large proportion of oil separated from the substrate and moving around in the bath. Thus, if the substrate is removed from the agitated bath and rinsed, it is freed of most of its soil burden.

This model for oily soil removal does not apply to systems where the oil soil becomes emulsified. Emulsified oil droplets are characterized by a surface layer that acts either as a physical barrier or as a potential energy barrier against coalescence, or against coming into intimate contact with the substrate. Such droplets must overcome this potential barrier to redeposit on the substrate, whereas the free-floating unemulsified oil droplets of the roll-up model redeposit on contact without surmounting any barrier, energetic or physical. The mathematical model for emulsification can be found in standard treatises on emulsification. It resembles the model for solid-soil dispersion outlined above, although there are many important differences of detail.

Solubilization. The role of micellar solubilization (as the term is used in the physical chemistry of surfactants) in oily-soil removal has been debated for many years. The amount of oily soil that could be present in a normal wash load could not all be removed and held in micellar solution by anionic surfactants. On the other hand, nonionic surfactants could do so, because of their greater solubilizing ability. High solubilizing power is definitely linked with good detergency (80). Thus, a very direct relationship between the solubilizing power of a surfactant for the test dyestuff Orange OT and its ability to remove polar solid from steel surfaces was established (81). In a practical detergency range lying between the surfactant concentration that gives 90% soil removal and a surfactant concentration twice that high, the relationship is expressed by the equation below, where D is the detergency value, S is the solubilization value, and K_1 and K_2 are constants:

$$D = K_1 S + K_2$$

Furthermore, in a series of polyoxyethylene nonylphenol nonionic surfactants, the value of K_1 varied linearly with the HLB number of the surfactant. The value of K_2 varied linearly with the log of the interfacial tension measured at the surfactant concentration that gives 90% soil removal. Carrying the correlations still further, it was found that from the detergency equation of a single surfactant with three different polar soils, K_1 was a function of the soil's dipole moment and K_2 a function of the soil's surface tension (81).

Detailed thermodynamic and mechanistic analyses of solubilization and related mechanisms are given in References 17 and 82–84. These works show that under proper circumstances, solubilization can make a significant contribution to oily-soil removal.

Phase Changes at the Soil–Bath Interface. Closely related to solubilization is a phenomenon that involves polar organic soils and surfactant solutions. If a complete phase diagram is plotted for a ternary system containing sodium dodecyl sulfate (or glycerol oleate), and water, several important and unusual features are noted. A large area represents a liquid phase consisting of a microemulsion, where the dispersed particles are so small that the system is isotropic, like the familiar soluble oils. Also, over another large area, a liquid crystalline phase is formed, containing all three components. This liquid crystalline phase flows like a liquid, at least in one direction. Flow perpendicular to the oriented planes is accomplished by folding the planes cylindrically, but the physical flow is still of the purely viscous type, with no yield point evident. These two phases, particularly the liquid–crystal phase, play an important part in detergency (85). Furthermore, liquid–crystal formation lowers interfacial tension (86). Although this phenomenon was demonstrated in tertiary oil recovery, the principles could also apply to oily-soil detergency.

When the polar organic component is a solid at ordinary temperatures, the addition of detergent and water markedly lowers the melting point; more specifically, as the temperature is raised, a point is reached where surfactant and water penetrate the solid. Thus, the ternary liquid–crystal phase might form spontaneously at room temperature by mixing the components, or, more precisely, an aqueous detergent solution can literally melt and liquefy a relatively large proportion of solid polar fatty matter (see LIQUID CRYSTALS). What actually happens when these two phases are placed together at a low temperature and slowly warmed is a slow interaction. At a definite critical temperature, however, penetration of the detergent solution into the polar material starts to take place rapidly and the mass soon becomes fluid (17). This interaction is especially favorable in the case of nonionic surfactants that are relatively soluble in polar organic compounds and are the active ingredients of choice when forming microemulsions. In addition, above certain temperatures they display multilayer adsorption isotherms (87–92).

These phenomena are most rapid and easiest to observe in fairly concentrated aqueous detergent solutions, that is, minimally 2–5% detergent solutions. In a practical qualitative way, this is a familiar effect, and there are many examples of the extraordinary solvency and cleaning power of concentrated detergent solutions, for example, in the case of fabric pretreatment with neat heavy-duty liquid detergents. Penetration can also be demonstrated at low detergent concentrations. As observed microscopically, the penetration occurs in a characteristic manner with the formation of a sheathlike structure, termed myelin; they are filled with isotropic liquid but have a liquid crystalline birefringent skin.

In a detersive system containing a dilute surfactant solution and a substrate bearing a solid polar soil, the first effect is adsorption of surfactant at the soil–bath interface. This adsorption is equivalent to the formation of a thin layer of relatively concentrated surfactant solution at the interface, which is continuously renewable and can penetrate the soil phase. Osmotic flow of water and the extrusion of myelin forms follows the penetration, with ultimate formation of an equilibrium phase. This equilibrium phase may be microemulsion rather than liquid crystalline, but in any event it is fluid and flushable

from the substrate surface. This phase change effect explains the deterative behavior of sucrose fatty esters in admixture with alkylarenesulfonates (93).

Measurement of Detergency

The measurement of detergency can be approached from two different points of view. The theoretical approach is concerned with the relative quantity of soil bound to the substrate before and after washing. In this case, measurement is a necessary analytical procedure in the study of the detergency mechanism. The second approach emphasizes the development of reproducible laboratory methods that predict the results of practical cleaning operations. In the development of new household-cleaning compositions, for example, the formulator must know whether his products outperform others under actual use conditions. Realism, or accuracy as it is usually termed, is a prime requisite of any detergency test. It means good correlation between laboratory evaluation and the results of field testing. The practical field evaluation is usually made on the basis of specifications, sometimes implied rather than explicitly expressed, that determine whether or not the cleaning results were satisfactory. In removing spinning lubricants from wool yarns, for example, the most satisfactory cleaning procedure is that which removes most soil from the substrate. In certain metal-cleaning operations, however, the satisfactory outcome is a piece of metal free of solid soil but carrying an even and easily perceivable layer of rust-preventive oil. In judging the cleanness of white fabrics, more interest is usually shown in fabric whiteness rather than in its actual soil content. Thus, most detergency evaluations in which white fabric is the substrate specify cleanness in terms of fabric whiteness or reflectance.

With these limitations in mind, the measurement of detergency in the laboratory requires the following components: A means for measuring or estimating the amount of soil on the substrate or the degree of cleanness both before and after washing; satisfactory substrates and soiling compositions; a means for applying soil to substrate in a realistic manner; and a realistic and reproducible cleaning device. These fundamental requirements apply regardless of the particular type of substrate that is being cleaned. More attention in this area has been centered on textile fabrics than on other substrates, but the various substrates and even human skin can be considered from the same point of view.

FABRIC DETERGENCY

Laundering. Reflectance is the most commonly used measurement for the whiteness of fabrics, although the transmittance of light by fabric specimens can also be used. The most commonly used instrument for reflectance measurement is the Gardner colorimeter, although the Zeiss Elrepho is also used. For general detergency, the grayness of the fabric is measured. Color effects can also be measured, and fabric yellowing is especially important. It is masked by fluorescent whitening agents (FWA). Special filters are available to eliminate this effect, and whitening caused by soil removal can be distinguished from that of FWA deposition.

As would be expected, no single artificial soil or combination of artificial soils can adequately model natural soils in the household laundry. Natural soils vary widely within a single household, between households, and between regions. In addition, natural soiling includes effects of aging and repetitive soiling that are difficult to simulate artificially. Consequently, there is no single set of test conditions that can consistently rate a group of detergents in a completely realistic manner. Limited correlations are usually obtained by suitably adjusting both the laboratory procedure and the practical full-scale procedure.

These problems can be dealt with by using artificial test cloths impregnated with various approximations of natural soils such as vacuum cleaner dust, dirt from air conditioner filters, clays, carbon black, fatty acids, dirty motor oil, and artificial sebum, either alone or in combination (37,94–98). The soils are applied by spraying, immersion, or padding. If the soils are carefully applied, reproducible results can be obtained. Soil test cloths can be of great help in detergency studies, when used with an understanding of their limitations.

The device most widely used for laboratory fabric detergency measurements is the Terg-O-Tometer (U.S. Testing Co.), a miniature agitator washer. Just as it is difficult to model natural soils, it is difficult to model a full-scale washing machine. The Terg-O-Tometer consists of four or six small agitator washers in 2-L beakers. Water and detergent are placed into the beakers, and the temperature controlled by a water bath. Standard-soil cloths are added and washed for 5–15 min. The speed as well as the angle of oscillation of the agitators is adjustable. The soil cloths are then rinsed, usually in the Terg-O-Tometer, dried, and the reflectances read. The performance of various detergent mixtures can be compared on the basis of the final reflectance R_f of the washed soil cloths. However, it is generally more useful to express the cleaning as percentage detergency, $\%D$ where R_i is the initial reflectance (before the wash) of the soil cloths and R_o the reflectance of the unsoiled fabric used to prepare the soil cloths, as shown in the equation below.

$$\%D = \frac{R_f}{R_o} \times \frac{R_i}{R_i} \times 100$$

Redeposition of soil can be estimated simultaneously with net soil removal (detergency) by including a white swatch with the soiled swatches. It also can be estimated separately by adding a measured amount of soil to a fresh wash bath, then treating the white swatch in this standard-soil bath. The former method is, superficially at least, more realistic. The latter method, however, is frequently of greater value in development work because both the quantity and composition of the soil are known and can be controlled.

Grayness of a fabric swatch is not directly proportional to its content of black pigment (or artificial soil). A basic formula relating reflectance to the pigment content or concentration can be applied to the evaluation of detergency test swatches (51,99–101). In simple form, an adaptation of the Kubelka-Munk equation, it states that the quantity $(1 - R)^2 / 2R$ (where R is the fraction of light reflected from the sample) is a linear function of the soil content of the sample.

In some cases, it may be impossible, or undesirable, to measure the amount of soil by reflectance. Soil can also be determined by extraction and weighing the cloths, or weighing the washed cloths.

A number of excellent studies have used a variety of radiolabeled soils to investigate the removal of small amounts of colorless soils such as oils (102–104). By proper use of different radiolabels (such as ³H and ¹⁴C), the preferential removal of various components in a soil mixture can be followed. In these cases in particular, detergency can also be calculated from measurements of the amount of radioactivity that is removed from the fabric and is found in the wash liquor.

The above tests all measure the ability of a washing system to remove soil from a fabric in a single wash. However, the detergent or washing machine is often judged by the way the linen appears after several soil-wash cycles. After a series of soilings and washings, the linen acquires an off-white gray or yellowish shade caused by soil accumulation and chemical changes in the soil. Although soil accumulation tests are more tedious than soil removal tests, they give the most realistic results. A laboratory-scale soil accumulation test using vacuum cleaner dirt and small swatches of fabric has been described (105).

Textile Mill Operations. Detergency is important in textile finishing because small quantities of foreign matter on the goods can interfere seriously with dyeing and other finishing treatments. Furthermore, the goods are expected to be uniformly and thoroughly clean when sold. Many detergency tests in this area are of the semipractical type, ie, test swatches are analyzed for soil content. This analysis generally consists of gravimetric determination of the soil content either directly from the fabric weight, or by extraction (see TEXTILES).

HARD SURFACE DETERGENCY

Despite the variety of hard-surface objects that are purposefully cleaned at regular intervals, detergency has been studied quantitatively in relatively few cases only. The small-scale user normally judges washing results as satisfactory or unsatisfactory. If satisfactory results are obtained with the amount of detergent and the degree of mechanical action employed, the user is not interested in minor qualitative differences. In those areas where specifications are important and where differences among detergents or mechanical washing equipment are readily perceivable, quantitative methods for measuring detergency have been developed.

In specific cases of metal cleaning where small amounts of residual soil must be detected and are difficult to measure by conventional means,

radiotracer methods have been employed (106). Interest in these techniques has been stimulated by the development of methods for decontaminating hard surfaces subjected to atomic fallout (107).

Quantitative measurements have been obtained for ceramics and glass, metals, and organic surfaces such as painted and plastic tile.

Glassware and Dishwashing. Dishes are washed either by hand or in an automatic dishwashing machine. Hand-dishwashing detergents are generally high foaming compositions containing organic surfactants as the main ingredient. The consumer judges efficiency not only by the cleanness of plates but also by foam persisting throughout the operation. Evaluation of hand-dishwashing products by manufacturers simulates this procedure. The number of plates that can be washed clean, judged visually without or with a color or fluorescence indicator, and the number of plates necessary to kill the foam in the dishpan is taken as the measure of deterative efficiency (108,109).

More objective laboratory methods employ a mechanical device such as a Terg-O-Tometer (110). Food soils are applied to microscope slides or glass tape rather than to actual plates. The soils are tagged with fluorescent materials or with dark pigment to facilitate measurement of residual soil. Reflectance or transmittance may also be read directly (111).

The foam stability of hand-dishwashing compositions can also be measured more directly and more quantitatively using mechanical means to whip up a foam and adding increments of food soil to a predetermined no-foam end point (112).

The detergents in automatic-dishwashing compositions are largely inorganic, but evaluation of residual soil is essentially the same as in hand-dishwashing tests. Low foam avoids inhibition of the free movement of the rotor. Clarity of glassware is a particularly prized performance feature of home dishwashers and special photometric methods have been reported for measuring freedom from water spots or filming (haze) (113). More frequently, however, the presence of spots and filming is assessed visually in a box in which light is beamed at the interior of inverted glasses in an otherwise black environment.

In restaurant operations, sterilization of dishes is an additional requirement (see **STERILIZATION TECHNIQUES**). Sterilization is determined by the usual swabbing and culturing methods or by employing bacteria tagged with radioactive phosphorus and counting residual radioactivity on the washed dishes (108).

Metal Cleaning. The purpose of cleaning steel is to remove dirt and leave the article in a state where it can be delivered for use without further finishing (see **METAL SURFACE TREATMENTS**). The surface must therefore be covered with a tenacious corrosion-resistant coating as it emerges from the cleaning bath. Many emulsion cleaners remove lubricants and other unwanted dirt while depositing an anticorrosive coating on the metal. The primary test for efficacy in this situation is a corrosive test of the cleaned article.

A more thorough cleaning prepares the surface for further finishing, eg, electroplating or painting, or application of an oxide finish. The basic objective in this case is to ensure maximum adhesion of the finish by freeing the surface of all foreign matter (114). Frequently, the cleaning compositions are straight organic solvents, hydrocarbons or chlorinated hydrocarbons, but in many cases aqueous detergent solutions are used. Regardless of the composition of the cleaning product, detergency test methods involve soiled test specimen, a standard method or device for performing the cleaning operation, and a quantitative method for estimating the extent of soil removal. The test specimens are mostly metal panels, or coupons, 10–300 cm² in area. They are carefully cleaned, usually by abrasive action, and then soiled. Soils may vary widely but often are based on lubricating oil or asphalt. A dark pigment or fluorescent marker is added if the cleaning result is to be evaluated visually or by photometry. A radioactive tracer, compatible or identical with the soil, may also be added for assessment of residual soil. The cleaning device is often a miniature version of large-scale cleaning equipment that provides soaking of the test article and agitation by hydraulic action or boiling.

When a quantitative estimate of residual soil is not called for and the suitability of a metal surface for further finishing needs to be assessed, the water-break test is used. The term water-break refers to the behavior of a water film on a smooth greasy surface. When the film becomes sufficiently thin by drainage, it suddenly breaks into islands or droplets between which the surface appears dry. On the other hand, when a film drains from a clean water-wettable, nongreasy surface, it becomes progressively thinner and finally disappears by evaporation without ever breaking into droplets. Such a surface is said to be free from water-break.

In a similar procedure, the atomizer test, which depends on the behavior of an advancing rather than a receding contact angle, a fine mist of water is applied to the metal surface and the spreading of water is observed. On a clean surface, water spreads to a uniform film. With oleic acid as the test soil, the atomizer test can detect the presence of 10⁻⁵ mg of soil per cm², less than a monomolecular layer (115). For steel that is to be electroplated, the copper dip test is often employed. Steel is dipped into a cupric salt solution and the evenness of the resulting metallic copper deposit is noted.

Correlation of all aspects of the test method with the practical system of interest is always important. The test used for dairy cleaning is an excellent example (116). Milk is used to tag the soil with radioactive ⁴⁵Ca by an exchange with radioactive CaCl₂. This treatment is applied to stainless steel planchets by suspending the planchets in milk under actual pasteurizing conditions.

Organic Surfaces. Tests for detergency on organic surfaces such as painted walls and plastic tile generally include a rubbing or sponging step corresponding to the manner in which such surfaces are cleaned in practice. The method adopted by the Chemical Specialties Manufacturers' Association includes scrubbing a painted glass surface soiled with a marking pencil in a Gardner Straight Line Washability Machine and assessing deterative efficacy, visually on a seven point scale (117). In a similar method, a preconditioned linoleum tile is coated with an oily pigment soil and scrubbed in a Gardner machine; the deterative efficacy is measured by reflectance change (118).

Detergent Manufacture

Liquid Products. The manufacture of liquid detergent products is generally a straightforward process requiring batch equipment with provisions for metered addition of individual ingredients, agitation, and if needed, heating and cooling. Capital cost can vary depending on the degree of automation.

Spray-Dried Products. The manufacture of powdered product is more complicated. High-pressure spray-drying of an aqueous slurry has replaced the earlier process in which a solidified cake of the product had to be broken up mechanically. Spray-drying equipment requires a relatively high capital outlay. The resulting product, however, is characterized by several desirable features: a high concentration of surfactant can be accommodated in the finished product; the product consists of hollow beads that dissolve readily in the washing solution; powders are considerably less dusty than those produced by the earlier process; powders generally do not lump and most compositions can be packaged in cartons without special liners; and bulk densities are obtainable within the general range of 0.25–0.65 g L.

The first stage in preparing spray-dried products involves producing a slurry of liquid and solid ingredients. Two processes are available, a batch process or a continuous process. The batch process involves preparing a slurry in a crutcher, which is a mixing vessel with heavy-duty agitation and provision for heating. Solid and liquid components are combined to form a homogeneous slurry. It is normal with a batch process that two crutchers are employed; while one crutcher batch is discharged to the spray-drying tower, another batch is prepared in the second crutcher. The automated batch process is preferred when there are reaction, hydration and crystallization processes involved.

The continuous process can offer shorter residence times and requires a high degree of automation (119). In addition, continuous slurry preparation can permit high solids concentrations and hence reduce the evaporative load in the spray-drying tower. The resulting energy savings can be significant in large installations.

Acids such as fatty acids and alkylbenzenesulfonic acids are neutralized with NaOH during slurry preparation to form soap and sodium alkylbenzenesulfonate, respectively.

After preparation, the slurry is transported to an aging vessel. During residence time of 20–30 min, the neutralization process, hydration of sodium tripolyphosphate and structural changes in the slurry are completed to provide a homogeneous composition. By means of a high pressure pump, the slurry is conveyed to the spray-drying tower under a pressure of ca 10 MPa (100 atm). A representative spray drying tower layout is shown in Figure 7.

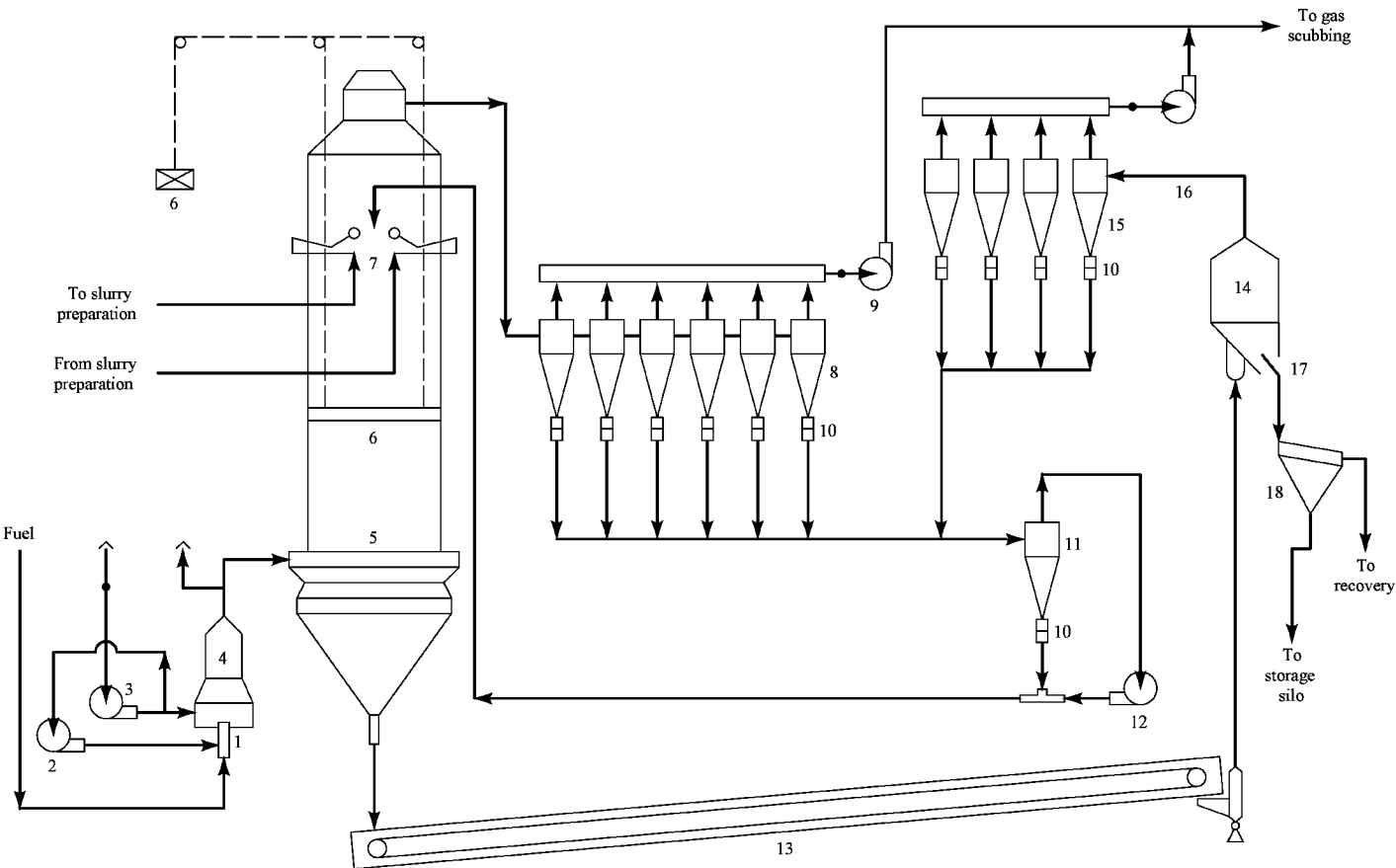


Fig. 7. Plant for spray-drying of detergents and soaps. 1, Burner; 2, air combustion; 3, cold air fan; 4, hot air generator; 5, spray-drying tower; 6, cleaning ring; 7, nozzles circuit; 8, tower suction cyclones group; 9, tower suction fan; 10, double-flap dischargers; 11, recovery powder cyclone; 12, recovery powder fan; 13, belt conveyor; 14, air lift; 15, air lift suction cyclones; 16, air lift suction fan; 17, double-flip dischargers; 18, vibrating sieve.

The slurry, at 80–100°C, is forced through nozzles of 2.5–3.5 mm dia arranged on a nozzle ring. In the tower, the slurry encounters hot air that has entered the tower at 250–350°C. Air flow is normally countercurrent. Air circulation is provided by two blowers. A third blower provides suction that carries fines into the collecting cyclones. From there the fines may be reintroduced into the upper part of the tower where they are contacted by wet slurry descending in the tower. Alternatively, the fines may be added to subsequent crutcher batches. Upon exit, the powder is conditioned during passage via a belt conveyor and an airlift to the packaging machinery.

The capacity of spray-drying towers is influenced by the formulation of the powder being produced. For large spray-drying towers, throughputs of up to 30 tons per hour can be expected.

Because of stringent air pollution rules the exit gases are wet-scrubbed in brine since high NaCl concentrations reduce foam formation. NaOH scrubs out SO₂ from sulfur-containing fuel. The scrubbing solution, saturated with detergent fines, is recycled to the water tank for slurry preparation. As shown in Figure 7, the tower is also provided with a cleaning ring that is moved along the tower walls.

Although spray-drying accommodates relatively high content of surfactants, certain types, such as the alkanolamides and some nonionic surfactants are best added to the product after spray-drying. Post addition not only protects the surfactant from the heat of the tower but also prevents the formation of aerosols in the exit gas. Aerosols are more difficult to trap in the scrubbing system than solid fines. They are formed by unsulfonated matter from the manufacture of LAS and nonionic surfactants with short ethylene oxide chains (120).

Dry-Blended Product. In addition to lower capital outlay, dry-blending requires considerably less processing energy. Final product density, which is usually near unity, depends on the density of the starting materials and the nature of equipment used to blend these materials. Modern mixing and blending equipment, if properly controlled, can give product density and particle sizes comparable to spray-dried products.

Agglomerated Products. The process of agglomeration is intermediate between spray-drying and dry-blending. Process water concentrations are between 35 to 40% in a crutcher slurry and essentially zero in dry blending. In agglomeration, a liquid is sprayed onto a continuously agitated powder. Equipment designs include stationary mixers, rotating mixers with spray nozzles, and rotating blenders with a liquid dispersion bar, either twin shell or continuous zigzag (121). Automatic dishwashing detergents are mostly manufactured by agglomeration (122). Liquid components of the formulation, for example silicate, detergent actives, and water, can be used as the agglomerating liquid. Other examples of agglomeration processes include the hydration of sodium tripolyphosphate, coloring (speckling) of detergent powders, and the agglomeration of spray tower fines.

Analysis

The literature on analytical methods is voluminous and not easily summarized (123–130). Often the greatest expertise in the analysis of complex detergent mixtures resides with in-house personnel in individual companies who may regard their methods as proprietary information.

An unknown commercial detergent may contain some combination of anionic, nonionic, cationic, and possibly amphoteric surfactants, inorganic builders and fillers as well as some minor additives. In general, the analytical scheme includes separation of nonsurfactant and inorganic components from the total mixture, classification of the surfactants, separation of individual surfactants, and quantitative determination (131).

Health and Safety Factors As a class, surfactants and detergent products are among the most widely used chemical compositions. Almost everyone is exposed to these products on a daily basis in situations that range from ingestion of food-grade emulsifiers to intimate contact of skin and eyes with personal-care and laundry products. Safety is therefore a matter of great importance (132,133). Ranges of surfactant LD₅₀ values are shown in Table 2.

Table 2. Rat Oral LD₅₀ Values of Surfactant^a

Type of compounds	Oral LD ₅₀ , mg kg	
alkylbenzenesulfonates	700–2,480	
alcohol ethoxylates	1,600 to greater than 25,000	
sulfated alcohol ethoxylates	7,000 to greater than 50,000	
alcohol sulfates	5,000–15,000	

^a Ref. 132.

Under conditions of normal use, detergent products are not hazardous to users. Nonetheless, surfactants possess some toxicity, and they are mild irritants. Particularly under conditions of misuse, such as accidental ingestion or spillage, they can produce irritation and discomfort in the form of nausea and vomiting, as well as irritation to skin and eyes. The long-term effects, however, are minimal (134).

In the last two decades (1980s and 1990s), governmental concern with the safety of chemicals has led to a number of legislative acts that regulate the manufacture and sale of chemicals including detergents. Even before this, the Food and Drug Act was passed in 1938. Food-grade emulsifiers fall under its provisions, as well as products containing antimicrobial agents such as deodorant soap bars.

Like other chemicals, new substances introduced into detergent products are regulated by the TSCA of 1977. Since its purpose is to prevent chemicals with long-term deleterious effects from entering the marketplace, the emphasis in testing is on long-term effects such as carcinogenicity, mutagenicity, and teratogenicity.

The manufacture of surfactants and of detergent products is regulated by OSHA. Dust concentration in detergent plants as well as factory noise levels are the primary areas of relevance, since the individual components in these products are essentially nonhazardous. Of more immediate concern to the detergent industry is the Federal Hazardous Substances Act (FHSA) of 1960 and the Consumer Products Safety Act (CPSA) of 1972. The FHSA defines specific labeling requirements, such as "Danger" for extremely flammable, corrosive, or highly toxic substances, and "Warning" or "Caution" for less hazardous materials.

To assess the degree of hazard, animal acute oral toxicity data are generally relied upon even though such data have limitations in their applicability for predicting possible human effects in the case of a specific exposure. Ranges of detergent LD₅₀ values ratings are given in Table 3.

Table 3. Detergent Animal Acute Oral LD₅₀ Versus Probable Lethal Dose for a Human Adult^a

Animal acute oral LD ₅₀ , g kg	Toxicity rating	Probable lethal oral dose for 70 kg person, g ^{b,c}
<0.005	6 super toxic	taste, <0.35
0.005–0.050	5 extremely toxic	0.35–4.9
0.05–0.50	4 very toxic	4.9–28.3
0.5–5.0	3 moderately toxic	28.3–453
5.0–15.0	2 slightly toxic	453–1000
>15	1 practically nontoxic	>1000

^a Ref. 135.

^b 0.35g = 7drops; 4.9g = 1tsp; 28.3g = 1oz.

^c To convert g to lb, multiply by 0.0022. 1 lb is approximately 1 pt.

The range of values for several representative categories of detergent products is given in Table 4.

Table 4. Acute Oral LD₅₀ Ranges for Detergent Product^a

Detergent type	Albino rat, oral LD ₅₀ , g kg
heavy-duty, laundry	
granular	2–7
liquid	2–9
hand-dishwashing, liquid	5–20
automatic dishwashing	2–7
floor and wall cleaner	
crystalline	4–6
liquid	8– >16
rug cleaner	7–9
laundry presoak with enzyme, granular	3–11
fabric softener	>10
toilet bar, soap, or synthetic base	7–20
shampoo	
plain	>10
medicated	3–10

^a Ref. 136.

Because of emesis, it is unlikely that appreciable quantities of most common soaps and detergents could be ingested accidentally.

Methods of testing for eye and skin irritation potential have been reviewed (137). The official FHSA procedure for evaluating ocular irritation potential of detergent products is a modified Draize rabbit eye test (138). Some controversy surrounds this method at present, and a search for a procedure less injurious to test animals is in progress. In general, the order of irritation is cationic > anionic > nonionic (139).

Skin irritation potential is assessed by patch tests. More serious than simple irritation is the potential of a product to cause sensitization, ie, to cause a subject to become allergic to even very small amounts of the product.

Several test methods have been devised for sensitization. In the Magnussen-Kligman procedure, guinea pigs are injected just below the skin with a slightly irritating dose of the test substance with and without an antigen. After some days, a patch test is taken on the injection site to stimulate the skin to react. After several weeks, another patch test is taken, this time on an untreated site. A positive reaction is an indication of the sensitizing potential of the test substance (140).

Photosensitization, the potential of a product to cause sensitization in the presence of sunlight, is similarly evaluated by taking a patch test on guinea pigs before and after exposure to uv radiation (141).

Controversy over test methodology, and concern with the welfare of test animals, has been highly publicized in the last decade, and various states have proposed legislation to ban animal tests. Significant effort has been devoted to developing nonanimal alternative tests and predictive methods. Progress has been made, but no entirely satisfactory substitute has been found to date (142).

Environmental Considerations

The introduction of surfactant products into the environment, after use by consumers or as part of waste disposed during manufacture, is regulated by the Clean Water Act, the Clean Air Act, and the Resource Conservation and Recovery Act. In this respect, surfactants are subject to the same regulations as chemicals in general. There are, however, two areas of specific relevance to surfactants and detergent products, ie, biodegradability and eutrophication.

As early as 1947, incidents of foam blankets on bodies of water were reported. This was at a time when detergents containing synthetic surfactants were in the process of displacing soap as the principal heavy-duty laundry product. It was suspected that they were the cause of foam formation since they were less readily degraded by the microorganisms present in water. In contrast to soap with a straight-chain hydrophobe, alkylbenzenesulfonate, ABS, the predominant surfactant at the time, contained a highly branched hydrophobe, eg, a propylene tetramer alkyl group. The U.S. surfactant industry, represented by The Soap and Detergent Association (SDA), took the lead in the investigation of the problem, sponsoring a large number of field studies in the development of test methods and analytical procedures for measuring the presence of surfactants in the environment. For anionic surfactants like ABS, the Methylene Blue Active Substance (MBAS) test, which depends on the formation of a chloroform-soluble complex of the anionic surfactant with cationic methylene blue, provides the first indication of the extent to which a surfactant has been degraded (143). Nonionic surfactants are assayed by complex formation with cobalt thiocyanate (144).

Among the methods for simulating the fate of the surfactant in the environment, the river die-away test (145) and the shake-flask method (146) have proved acceptable for fast screening and routine use. The semicontinuous activated sludge (SCAS) method is more time-consuming, but is more accurate and reproducible (147). Determination of biological oxygen demand (BOD) also provides useful data on biodegradation (148).

The extensive investigations led to the conclusion that a branched hydrophobe impedes the rate and extent of degradation of surfactants by microorganisms. The most immediately apparent remedy, therefore, was to replace the propylene tetramer in ABS with a straight hydrocarbon chain giving straight-chain alkylbenzenesulfonate, so-called linear alkanesulfonate (LAS). At the same time, commercialization of the Ziegler process for the oligomerization of ethylene provided another route to straight-chain hydrophobes that could easily be converted to detergent alcohols and straight-chain nonionic surfactants. By 1965, the U.S. detergent industry had completed a voluntary switch from hard to soft surfactants at a cost that has been estimated at ca \$150 × 10⁶. In addition to ABS, other surfactants based on propylene oligomers, such as alkylphenol derivatives, have largely disappeared from U.S. consumer laundry products.

Even though the biodegradability problem has been solved for all practical purposes, the subject continues to receive considerable attention. The biodegradation of LAS has been studied intensively, and several mechanistic pathways have been identified such as β- and ω-oxidation as well as reductive and oxidative desulfonation (149). Investigation of the biodegradation of LAS, alcohol ethoxylates, and alkylphenol ethoxylates in the laboratory and under sewage plant operating conditions showed that LAS and straight-chain alcohol ethoxylates and their sulfates degrade to CO₂ and H₂O (150,151).

Eutrophication. This term, which denotes excessive nutrition or overfertilization, has been applied to the contribution excessive amounts of phosphorus may make to the growth of algae under certain conditions. Phosphorus in water supply originates from run-off of agricultural fertilizers, human excrement, and sodium tripolyphosphate present in detergent formulations. It has been estimated that 25–30% of the phosphorus in waste water comes from laundry detergents, and that detergents contribute about 3% of the phosphorus annually entering U.S. surface waters (152). Excessive algal growth in stagnant bodies of water contributes to oxygen depletion, which causes starvation of marine life and eventually leads to the death of lakes through silting. Phosphorus can be removed from the effluent of sewage treatment facilities by treatment with materials such as alum. Expansion of sewage treatment has been the method of choice in some areas of the world, notably Sweden, to assure acceptably low phosphorus content in environmental waters (153). In the United States, and later in Western Europe, detergent phosphates were singled out in the early 1960s as the cause of eutrophication, and their removal from consumer laundry formulations was proposed as a feasible approach to improvement of environmental water quality. Many states and a number of local jurisdictions have banned detergent products containing phosphate.

The efforts of the detergent industry toward solution of its part of the eutrophication problem are, at this point, less complete than its response to the biodegradability problem. Soda ash, Na₂CO₃, sodium silicate, and, to a lesser extent, sodium citrate formed the basis of the early formulations marketed in the areas where phosphates were banned. Technically, these substances are considerably less effective than sodium tripolyphosphate. As a precipitant builder, soda ash can lead to undesirable deposits of calcium carbonate on textiles and on washing machines.

At the same time, the industry embarked on an intensive search for phosphate substitutes. Of a very large number of experimental organic builders, a few substances reached commercialization or near-commercialization including trisodium nitrilotriacetate (NTA), trisodium carboxymethoxysuccinate (CMOS) (154) and trisodium carboxymethyltartronate (155). As discussed above, sodium citrate ether carboxylates have achieved widespread use as phosphate substitutes. Polymeric builders (polyelectrolytes) proved to be effective calcium sequestrants, but failed to satisfy the criterion of acceptable biodegradability. Interestingly, some monomeric polycarboxylates proved to be even more powerful calcium sequestrants than sodium tripolyphosphate but were not sufficiently biodegradable (156).

Trisodium nitrilotriacetate, a sequestrant of effectiveness comparable to sodium tripolyphosphate, reached commercialization in the late 1960s. However, because some laboratory findings suggested potential teratogenicity, it was withdrawn from the market in 1970 at the request of the U.S. Surgeon General. It has continued to be used in Canada and elsewhere. On the basis of many studies supporting the safety of NTA, the EPA dropped its opposition to NTA and this builder, in combination with type-A zeolite, briefly appeared in some nonphosphate consumer laundry products. However, continuing opposition to NTA on the state level resulted in its withdrawal from detergents and its reappearance seems unlikely at this time.

Laboratory assessment of the eutrophication potential of an experimental substance is less clear-cut than that of biodegradation. A frequently used method is the algal assay procedure in which a variety of algal cultures is grown in open shake flasks and the effect of test material on their growth is determined (152). In a second procedure, the MAAP test, one observes the behavior of a microcosm, consisting of bacteria, algae, zooplankton, sediment, and water taken from an oligotrophic lake. The diversity index, an indication of the number of algal species present, is a measure of the nutritive index of the system. A reduction in the diversity index is taken as an indication of the eutrophic potential of the test substance (158).

BIBLIOGRAPHY

" Detergency " in *ECT* 1st ed., Vol. 4, pp. 938–960, by A. M. Schwartz, Harris Research Laboratories; " Detergents " in *ECT* 1st ed., Suppl. 1, pp. 190–223, by A. M. Schwartz, Harris Research Laboratories; " Detergency " in *ECT* 2nd ed., Vol. 6, pp. 853–895, by A. M. Schwartz, Harris Research Laboratories; " Surfactants and Detersive Systems " in *ECT* 3rd ed., Vol. 22, pp. 332–432, by A. Cahn, Consultant, and J. L. Lynn, Jr., Lever Brothers Co.
1. C. B. Brown, *Research* (London) **1**(10), 46 (Oct. 1947); W. C. Powe and W. L. Marple, *J. Am. Oil Chem. Soc.* **37**, 136 (1960).
2. E. Kissa, in W. G. Cutler and E. Kissa eds., *Detergency: Theory and Technology*, Marcel Dekker Inc., New York, 1977, Chapt. 1.
3. W. W. Morgenthaler, *Inform* **2**, 6 (1991).
4. **U.S. Pat. 3,862,058** (Jan. 21, 1975), J. P. Nirschl and R. A. Gloss (to Procter & Gamble Co.).
5. **U.S. Pat. 3,936,537** (Feb. 3, 1976), R. J. Baskerville, Jr., and F. G. Schiro (to Procter & Gamble Co.).
6. **U.S. Pat. 4,062,647** (Dec. 13, 1977), T. D. Storm and J. P. Nirschl (to Procter & Gamble Co.).
7. **U.S. Pat. 4,141,841** (Feb. 27, 1979), S. R. McDonald (to Procter & Gamble Co.).
8. G. K. Greminger, Jr. and A. S. Teot, *J. Am. Oil Chem. Soc.* **55**, 122 (1978).
9. A. E. Siegrist, *J. Am. Oil Chem. Soc.* **55**, 114 (1978).
10. P. N. Christensen, P. Holm, and B. Smonder, *J. Am. Oil Chem. Soc.* **55**, 109 (1978).
11. C. A. Starace, *J. Am. Oil Chem. Soc.* **60**, 1025 (1983).
12. G. Jensen, *Tenside Surf. Det.* **27**, 30 (1990).
13. A. Gillis, *J. Am. Oil Chem. Soc.* **66**, 734 (1989).
14. A. P. James, " Chemistry of Peroxygen Bleaching, " paper presented at *83rd American Oil Chemists' Society*, Toronto, Canada, May 1992, Interox America Reprints, Houston, Tex., 1992.

15. U.S. Pat. **4,412,934** (Nov. 1, 1983), S. Y. Chung and G. L. Spadini (to Procter & Gamble Co.).

16. U.S. Pat. **4,663,071** (May 5, 1987), R. D. Bush, D. S. Connor, S. W. Heinzman, and L. N. Mackey (to Procter & Gamble Co.).

17. M. E. Ginn and J. C. Harris, *J. Am. Oil Chem. Soc.* **38**, 605 (1961).

18. J. A. Shaeiwitz and co-workers, *J. Colloid Interface Sci.* **84**, 47 (1981).

19. K. W. Dillan, E. D. Goddard, and D. A. McKenzie, *J. Am. Oil Chem. Soc.* **56**, 59 (1979).

20. *Ibid.*, **57**, 230 (1980).

21. H. L. Benson and K. H. Raney, *Shell Technical Bulletin, SC: 1182-90*, Shell Chemical Co., Houston, Tex., 1990.

22. E. E. Dreger and co-workers, *Ind. Eng. Chem.* **36**, 610 (1944).

23. H. Stupel, *Synthetische Wasch-und Reinigungsmittel*, Konradin-Verlag, Stuttgart, 1954.

24. G. E. Barker, *J. Am. Oil Chem. Soc.* **28**, 317 (1951).

25. H. R. Suter and M. G. Kramer, *Soap Sanit. Chem.* **27**(8), 33 (1951).

26. R. L. Mayhew and R. C. Hyatt, *J. Am. Oil Chem. Soc.* **29**, 357 (1952).

27. J. V. Karabinos, G. E. Bartels, and G. E. Kapella, *J. Am. Oil Chem. Soc.* **31**, 419 (1954).

28. F. N. Baumgartner, *Ind. Eng. Chem.* **46**, 1349 (1954).

29. M. S. Furry and E. M. O'Brien, *Am. Dyestuff Re.* **41**, 861 (1952).

30. R. Bernstein and H. Sosson, *Soap Sanit. Chem.* **30**(1), 46 (1954).

31. F. W. Gray, J. F. Gerecht, and I. J. Krems, *J. Org. Chem.* **20**, 511 (1955).

32. H. A. Ludeman, J. A. Balog, and J. C. Sherrill, *J. Am. Oil Chem. Soc.* **35**, 5 (1958).

33. T. G. Jones, in K. Durham, ed., *Surface Activity and Detergency*, Macmillan and Co., Ltd., London, 1961.

34. T. G. Jones and J. P. Parke, *Proceedings of the 4th International Congress on Surface Activity*, 3rd ed., University of Mainz Press, Germany, 1960, p. 178.

35. K. L. Stellner and J. F. Scamehorn, *J. Am. Oil Chem. Soc.* **63**, 566 (1986).

36. K. L. Stellner and J. F. Scamehorn, *Langmuir* **5**, 79 (1989).

37. B. J. Rutkowski, in W. G. Cutler and R. C. Davis, eds., *Surfactant Science Series*, Vol. 5, part 3, Marcel Dekker, Inc., New York, 1981, p. 1013.

38. S. Stawitz, W. Klaus, and P. Hoepfner, *Fette Seifen Anstrichm.* **60**, 94 (1958).

39. J. W. Hensley and C. G. Inks, *Textile Res. J.* **29**, 505 (June 1959).

40. N. K. Adam, *J. Soc. Dyers Colour.* **53**, 121 (1937).

41. W. Kling, E. Langer, and I. Haussner, *Melliand Textilber.* **25**, 198 (1945).

42. J. C. Stewart and C. S. Whewell, *Textile Res. J.* **30**, 903 and 912 (Dec. 1960).

43. J. Compton and W. J. Hart, *Textile Res. J.* **23**(3), 158 and 418 (1953); **24**(6), 263 (1954).

44. W. Kling and H. Mahl, *Melliand Textilber.* **35**, 640 (1954).

45. T. H. Shuttleworth and T. G. Jones, *International Congress on Surface Activity*, London, vol. 4, 2nd ed., 1957, Academic Press, Inc., New York, pp. 52-59.

46. V. W. Tripp and co-workers, *Textile Res. J.* **28** (June 1958).

47. J. Berch and H. Peper, *Textile Res. J.* **33**, 137 (Feb. 1963); J. Berch, H. Peper, and G. L. Drake, Jr., *Textile Res. J.* **34**, 29 (Jan. 1964).

48. W. C. Powe, *J. Am. Oil Chem. Soc.* **40**, 290 (July 1963).

49. J. L. Lynn, Jr., *Detergents In Depth '89*, Soap and Detergent Association, New York, 1989, p. 11.

50. J. C. Dickinson and R. C. Palmer, *J. Textile Inst.* **42**, T510 (Dec. 1951).

51. T. H. Vaughn, A. Vittone, Jr., and L. R. Bacon, *Ind. Eng. Chem.* **33**, 1012 (1941).

52. W. P. Utermohlen, Jr. and E. L. Wallace, *Textile Res. J.* **17**, 676 (1947); L. Loeb, P. B. Sanford, and S. D. Cochran, *J. Am. Oil Chem. Soc.* **41**, 120 (Feb. 1964).

53. H. Schott, *J. Am. Oil Chem. Soc.* **52**, 225 (1974).

54. D. G. Stevenson, *J. Soc. Dyers Colour.* **68**, 57 (1952).

55. I. J. Gruntfest and E. M. Young, *J. Am. Oil Chem. Soc.* **26**, 236 (1949).

56. I. Reich and R. D. Vold, *J. Phys. Chem.* **63**, 1497 (1959).

57. J. C. Harris, *Textile Research J.* **18**, 669 (Nov. 1948); J. C. Harris, *Soap Chem. Spec.* **34**(11), 40 (Nov. 1958); **34**(12), 59 (Dec. 1958); **35**(2), 57 (Feb. 1959).

58. A. S. Weatherburn, G. R. F. Rose, and C. H. Bayley, *Can. J. Res.* **27F**, 179 (1949); **27**, 426 (1949).

59. A. S. Weatherburn and C. H. Bayley, *Textile. Res. J.* **22**, 797 (1952).

60. K. W. Gardiner and L. B. Smith, *J. Am. Oil Chem. Soc.* **26**, 194 (1949).

61. L. H. Flett, L. F. Hoyt, and J. Walter, *Am. Dyestuff Rep.* **41**, 139 (1952); W. A. Fessler, *Am. Soc. Test. Mater. Spec. Tech. Publ.* **115**, 9 (1951).

62. A. L. Meader, Jr. and B. A. Fries, *Ind. Eng. Chem.* **44**, 1636 (1952).

63. K. McLaren, *J. Soc. Dyers Colour.* **66**, 521 (1950).

64. J. M. Lambert, *Ind. Eng. Chem.* **42**, 1394 (1950).

65. J. Tuzson and B. A. Short, *Textile Res. J.* **30**, 983 (Dec. 1960).

66. B. A. Short in Ref. 37, pp. 237-238.

67. B. Derjaguin, *Discuss. Faraday Soc.* **18**, 85 (1954); B. Derjaguin and L. Landau, *Acta Physicochim.* USSR **14**, 633 (1941).

68. E. J. W. Verwey and J. T. G. Overbeek, *Theory of the Stability of Lyophobic Colloids*, Elsevier Publishing Co., Amsterdam, 1948.

69. J. C. Harris, *Textile Res. J.* **28**, 912 (Nov. 1958).

70. K. Durham, *J. Appl. Chem.* **6**, 153 (1956); K. Durham in Ref. 45, p. 60.

71. H. Lange, *Kolloid-Z.* **154**, 103 (1957); **156**, 108 (1958); *Am. Dyest. Rep.* **50**, (12), 25 (June 12, 1961).

72. J. Ross, P. T. Vitale, and A. M. Schwartz, *J. Am. Oil Chem. Soc.* **32**, 200 (1955).

73. P. T. Vitale, *J. Am. Oil Chem. Soc.* **31**, 341 (1954).

74. A. S. Porter in Ref. 43, p. 103.

75. W. J. Schwarz, A. R. Martin, and R. C. Davis, *Textile Res. J.* **32**, 1 (Jan. 1962).

76. B. J. Rutkowski and A. R. Martin, *Textile Res. J.* **31**, 892 (Oct. 1961).

77. A. Bierman, *J. Colloid Sci.* **10**, 231 (1955).

78. O. F. Devereux and P. L. De Bruyn, *Interaction of Plane Parallel Double Layers*, MIT Press, Cambridge, Mass., 1963.

79. H. Lange, *J. Phys. Chem.* **64**, 538 (May 1960); *Kolloid-Z.* **169**, 124 (1960).

80. G. S. Hattiangdi, W. W. Walton, and J. I. Hoffman, *J. Res. Natl. Bur. Stand.* **42**, 361 (1949).

81. A. M. Mankowich, *J. Am. Oil Chem. Soc.* **39**, 206 (Apr. 1962).

82. R. Mallikarjun and D. B. Dadyburjor, *J. Colloid Int. Sci.* **84**, 73 (1981).

83. H. L. Rosano and M. Clausse, *Microemulsion Systems*, Marcel Dekker, Inc., New York, 1987.

84. R. Zana, *Surfactant Solution*, Marcel Dekker, Inc., New York, 1977.

85. A. S. C. Lawrence, *Nature* **183**, 1491 (1959).

86. E. I. Franses and co-workers, *J. Phys. Chem.* **87A**, 1547 (1980).

87. J. M. Corkill, J. F. Goodman, and J. R. Tate, *Trans. Faraday Soc.* **62**, 979 (1966).

88. N. A. Klimenko and A. M. Koganovskii, *Kolloidn. Zh.* **36**, 135 (1974).

89. N. A. Klimenko, A. A. Tryasorokova, and A. A. Permilovskaya, *Kolloidn. Zh.* **36**, 624 (1974).

90. N. A. Klimenko, A. A. Permilovskaya, and A. M. Koganovskii, *Kolloidn. Zh.* **36**, 728 (1974).
91. *Ibid.*, **37**, 870 (1975).
92. N. A. Klimenko and co-workers, *Kolloidn. Zh.* **37**, 873 (1975).
93. L. I. Osipow and F. D. Snell, *J. Am. Oil Chem. Soc.* **38**, 184 (1961).
94. F. L. Diehl and J. B. Crowe, *J. Am. Oil Chem. Soc.* **31**, 404 (1954); E. Goette, *Melliand Textilber.* **34**, 534 (1954).
95. C. Z. Draves and O. L. Sherburne, *Am. Dyestuff Rep.* **39**, P771 (1950).
96. H. L. Sanders and J. L. Lambert, *J. Am. Oil Chem. Soc.* **27**, 153 (1950).
97. R. E. Wagg in Ref. 43, pp. 35–42.
98. A. R. Martin and R. C. Davis, *Soap Chem. Spec.* **36**(4), 49 (Apr. 1960); (5), 73 (May 1960).
99. O. C. Bacon and J. E. Smith, *Ind. Eng. Chem.* **40**, 2361 (1948).
100. J. C. Harris, M. R. Sullivan, and L. E. Weeks, *Ind. Eng. Chem.* **46**, 1942 (1954).
101. I. Reich, F. D. Snell, and L. I. Osipow, *Ind. Eng. Chem.* **45**, 137 (1953).
102. B. E. Gordon and E. L. Bustin, *J. Am. Oil Chem. Soc.* **45**, 754 (1968).
103. B. E. Gordon and W. T. Shebs, *J. Am. Oil Chem. Soc.* **46**, 537 (1969).
104. J. C. Illman and co-workers, *J. Am. Oil Chem. Soc.* **47**, 379 (1970).
105. A. M. Schwartz and J. Berch, *Soap Chem. Spec.* **39**(5), 78 (May 1963).
106. E. H. Armbruster and G. M. Ridenour, *Soap Chem. Spec.* **31**(7), 47 (1955).
107. G. Segura and co-workers, *Chem. Ind. (London)*, 1270 (1953).
108. W. G. Mizuno, F. T. Lanners, and J. L. Wilson, *J. Am. Oil Chem. Soc.* **32**, 437 (1955).
109. A. Segesser and H. Stupel, *Seifen Oele Fette Wachse* **80**, 686 (1954).
110. G. R. Edwards and H. Stupel, paper presented at the 4th International Congress on Surface-Active Agents, Brussels, Sept. 1964, *Technical Bulletin SC:365–80*, Shell Chemical Co., Houston, Tex., 1980.
111. L. K. Bradacs and W. Schetelig, *J. Am. Oil Chem. Soc.* **41**, 161 (1964).
112. L. E. Weeks, J. C. Harris, and E. L. Brown, *J. Am. Oil Chem. Soc.* **31**, 254 (1954).
113. A. L. Kimmel, H. M. Gadbury, and D. O. Darby, *Soap Chem. Spec.* **37**, 51 (1961).
114. J. C. Harris, W. Stericker, and S. Spring, *ASTM Bull.* **204**, 31 (1955).
115. E. A. Saubestre, in M. J. Schick, ed., *Surfactant Science*, Vol. 1, Marcel Dekker, Inc., New York, 1967, Chapt. 19.
116. F. H. Firsching and H. E. Everson, *J. Am. Oil Chem. Soc.* **34**, 547 (1957).
117. *CSMA Designation DCC-04*, Chemical Specialties Manufacturers' Association, Washington, D.C., July 1973.
118. R. R. Adler, R. B. Albin, and B. M. Finger, *J. Am. Oil Chem. Soc.* **46**, 520 (1969).
119. A. Davidsohn, *J. Am. Oil Chem. Soc.* **55**, 134 (1978).
120. J. Kretschmann, H. Latka, and H. Reuter, *Waschmittelchemie*, Henkel & Cie, Heidelberg, 1976, pp. 235–252.
121. S. A. Kutli, *J. Am. Oil Chem. Soc.* **55**, 141 (1978).
122. R. J. Fuchs, J. M. Polkowski, and P. P. Carfagno, *Chem. Times Trends*, 37 (1977).
123. F. G. Longman, *The Analysis of Detergents and Detergent Products*, John Wiley & Sons, Inc., New York, 1975.
124. M. J. Rosen and H. Goldsmith, *Systematic Analysis of Surface-Active Agents*, 2nd ed., Wiley-Interscience, New York, 1972.
125. J. Kross, in M. J. Schick, ed., *Surfactant Science Series*, Vol. 8, Marcel Dekker, Inc., New York, 1975.
126. J. G. Karabinos, G. E. Kapella, and G. E. Bartels, *Soap Chem. Specialties* **6**, 30 and 41 (1954).
127. K. Burger, *Z. Anal. Chem.* **196**, 22 (1963).
128. L. E. Weeks, M. E. Ginn, and G. E. Baker, *Proc. Chem. Specialties Mfrs. Assoc.*, 150 (May 1957).
129. *Standard Methods for the Examination of Water and Wastewater*, 14th ed., American Public Health Association, New York, 1975.
130. R. A. Llenado, *Household Personal Products Industry*, **18**, 69, 71, 73 (1981).
131. H. G. Nadeau and S. Siggia in Ref. 115, Chapt. 24, pp. 422–431.
132. *Human Safety and Environmental Aspects of Major Surfactants*, NTIS document PB 301193, Soap and Detergent Association, Arthur D. Little, Inc., Cambridge, Mass., 1977.
133. Ref. 132, supplement, NTIS document PB 1182452, 1981.
134. A. R. Temple, *Detergents: In Depth '78*, Soap and Detergent Association, New York, 1978, pp. 36–41.
135. *Cleaning Products and Their Accidental Ingestion*, Scientific and Technical Report No. 5R, Soap and Detergent Association, New York, 1980, p. 9.
136. *Ibid.*, p. 10.
137. F. N. Marzulli and H. I. Maibach, *Advances in Modern Toxicology*, Vol. 4, John Wiley & Sons, Inc., New York, 1977.
138. *Food, Drug, and Cosmetic Law Reports* 233, 440, and 476.
139. Ref. 137, p. 440.
140. A. H. Gilbert, *J. Am. Oil Chem. Soc.* **58**, 66A (1981).
141. Ref. 137, Chapt. 14.
142. H. North-Root, Ref. 49, p. 32.
143. J. Longwell and W. D. Maniece, *Analyst* **80**, 685 (1955).
144. S. L. Boyer and co-workers, *Environ. Sci. Technol.* **11**, 1167 (1977).
145. P. J. Weaver, *Soap Chem. Spec.* **41**, 45 (1965).
146. Subcommittee on Biodegradation Test Methods, *J. Am. Oil Chem. Soc.* **42**, 989 (1965).
147. *Ibid.*, 990 (1965).
148. C. A. Vath, *Soap Chem. Spec.* **40**(2), 56 and 182 (1964); **40**(3), 55 and 108 (1964).
149. R. D. Swisher, *Surfactant Science Series*, Vol. 3, Marcel Dekker, Inc., New York, 1970.
150. R. J. Larson and G. E. Wentler, *Soap Cosmet. Chem. Spec.* **58**(5), 33 (1982).
151. R. L. Huddleston and A. M. Nielson, *Soap Cosmet. Chem. Spec.* **55**, 34, 36, 37, 44 (1979).
152. *C and E News*, **70**, 36 (1992).
153. H. O. Bouveng, *J. Am. Oil Chem. Soc.* **55**, 178 (1978).
154. **U.S. Pat. 3,635,830** (Jan. 18, 1972), V. Lamberti and M. Konart (to Lever Brothers Co.).
155. M. M. Crutchfield, *J. Am. Oil Chem. Soc.* **55**, 58 (1978).
156. H. C. Kemper and co-workers, *Tenside* **12**, 47 (1975).
157. *Algal Assay Procedure: Bottle Test*, National Eutrophication Research Program, Environmental Protection Agency, Corvallis, Ore., 1971.
158. D. Mitchell and J. C. Buzzell, Jr., *J. Sanit. Eng.* (8333), 453–465 (1971).

General References

A. W. Adamson, *Physical Chemistry of Surfaces*, 5th ed., John Wiley & Sons, Inc. New York, 1990.
M. J. Rosen, *Surfactants and Interfacial Phenomena*, 2nd ed., Wiley-Interscience, New York, 1989.
McCutcheon's Emulsifiers & Detergents, North Am. & International Ed., Glen Rock, N.J., 1991.
McCutcheon's Functional Materials, North Am. & International Ed., Glen Rock, N.J., 1991.

Detergents: In Depth, '89, Soap and Detergent Assoc., New York, 1989.

A Handbook of Industry Terms, Soap and Detergent Assoc., New York, 1981.

B. M. Milwidsky and D. M. Gabriel, *Detergent Analysis*, Halsted Press, New York, 1982.

A. R. Baldwin, ed., *Proceedings of the Second World Conference on Detergents: Looking Toward the 90's*, Am. Oil Chem. Soc., Chicago, 1987.

CSMA Detergents Division, *Test Methods Compendium*, 2nd ed., CSMA, Inc., Washington, D.C., 1985.

D. Myers *Surfactant Science & Technology*, VCH, New York, 1988.

J. Falbe, ed., *Surfactants in Consumer Products: Theory, Technology, and Application*, Springer-Verlag, New York, 1987.

T. F. Tadros, *Surfactants*, Academic Press, London, 1984.

M. J. Rosen, ed., *Structure/Performance Relationships in Surfactants*, Am. Chem. Soc. Washington, D.C., 1984.

K. L. Mittal and B. Lindman, eds., *Surfactants in Solution*, Plenum Press, New York, 1984.

A. M. Schwartz, J. W. Perry, and J. Berch, *Surface Active Agents and Detergents*, Vol. 2, Interscience Publishers, Inc., New York, 1958.

D. J. Shaw, *Introduction to Colloid and Surface Chemistry*, 3rd ed., Butterworth, London, 1981.

K. L. Mittal, ed., *Colloidal Dispersions and Micellar Behavior, ACS Symposium Series 9*, Am. Chem. Soc., Washington, D.C., 1975.

K. L. Mittal, ed., *Solution Chemistry of Surfactants*, Vol. 2, Plenum Press, New York, 1979.

P. Hiemenz, *Principles of Colloid and Surface Chemistry*, Marcel Dekker, Inc., New York, 1977.

R. L. Mittal, ed., *Micellization, Solubilization, and Microemulsions*, Vols. 1 and 2, Plenum Press, New York, 1977.

P. Becher, *Emulsions, Theory and Practice*, 2nd ed., Reinhold Publishing Corp., New York, 1977.

A. M. Schwartz and J. W. Perry, *Surface Active Agents: Their Chemistry and Technology*, Krieger, Huntington, N.Y., 1978.

A. Huthig, *Waschmittelchemie*, Henkel & Cie, Heidelberg, 1978.

M. Balsam and E. Sagarin, eds., *Cosmetics, Science and Technology*, Vols. 1, 2, 3, 2nd ed., Wiley-Interscience, New York, 1972, 1974.

M. J. Schick, ed., *Nonionic Surfactants, Surfactant Science Series*, Vol. 1, Marcel Dekker, Inc., New York, 1966.

K. Shinoda, ed., *Solvent Properties of Surfactant Solutions, Surfactant Science Series*, Vol. 2, Marcel Dekker, Inc., New York, 1967.

R. D. Swisher, ed., *Surfactant Biodegradation, Surfactant Science Series*, Vol. 3, Marcel Dekker, Inc., New York, 1970.

E. Jungermann, ed., *Cationic Surfactants, Surfactant Science Series*, Vol. 4, Marcel Dekker, Inc., New York, 1970.

W. Cutler and R. C. Davis, eds., *Detergency: Theory and Test Methods, Surfactant Science Series*, Vol. 5, Marcel Dekker, Inc., New York, 1972.

W. M. Linfield, ed., *Anionic Surfactants, Surfactant Science Series*, Vol. 7, Marcel Dekker, Inc., New York, 1976.

J. Cross, ed., *Anionic Surfactants: Chemical Analysis, Surfactant Science Series*, Vol. 8, Marcel Dekker, Inc., New York, 1977.

E. H. Lucassen-Reynders, ed., *Anionic Surfactants: Physical Chemistry of Surfactant Actions, Surfactant Science Series*, Vol. 11, Marcel Dekker, Inc., Basel, 1981.

B. R. Bluestein, ed., *Amphoteric Surfactants, Surfactant Science Series*, Vol. 12, Marcel Dekker, Inc., New York, 1982.

A. Kitahara and A. Watanabe, eds., *Electrical Phenomena at Interfaces Fundamentals, Measurements, and Applications, Surfactant Science Series*, Vol. 15, Marcel Dekker, Inc., New York, 1984.

C. A. Miller and P. Neogi, eds., *Interfacial Phenomena: Equilibrium and Dynamic Effects, Surfactant Science Series*, Vol. 17, Marcel Dekker, Inc., New York, 1985.

R. D. Swisher, ed., *Surfactant Biodegradation, Surfactant Science Series*, Vol. 18, Marcel Dekker, Inc., New York, 1987.

J. Cross, ed., *Nonionic Surfactants: Chemical Analysis, Surfactant Science Series*, Vol. 19, Marcel Dekker, Inc., New York, 1987.

W. G. Cutler and E. Kissa, eds., *Detergency: Theory and Technology, Surfactant Science Series*, Vol. 20, Marcel Dekker, Inc., New York, 1987.

R. Zana, ed., *Surfactant Solutions: New Methods of Investigation, Surfactant Science Series*, Vol. 22, Marcel Dekker, Inc., New York, 1986.

M. J. Schick, ed., *Nonionic Surfactants: Physical Chemistry, Surfactant Science Series*, Vol. 23, Marcel Dekker, Inc., New York, 1987.

J. M. Richmond, ed., *Cationic Surfactants: Organic Chemistry, Surfactant Science Series*, Vol. 34, Marcel Dekker, Inc., New York, 1990.

D. N. Rubingh, ed., *Cationic Surfactants: Physical Chemistry, Surfactant Science Series*, Vol. 37, Marcel Dekker, Inc., New York, 1991.

T. M. Schmitt, ed., *Analysis of Surfactants, Surfactant Science Series*, Vol. 40, Marcel Dekker, Inc., New York, 1992.

I. Piirma, ed., *Polymeric Surfactants, Surfactant Science Series*, Vol. 42, Marcel Dekker, Inc., New York, 1992.

G. Gloxhuber and K. Kunstler, eds., *Anionic Surfactants: Biochemistry, Toxicology, Dermatology, Surfactant Science Series*, Vol. 43, Marcel Dekker, Inc., New York, 1992.

A. M. Schwartz, in E. Matijevic, ed., *Surface and Colloid Science*, Vol. 5, John Wiley & Sons, Inc., New York, 1972.

B. J. Carroll, in E. Matijevic, ed., *Surface and Colloid Science*, Vol. 9, John Wiley & Sons, Inc., New York, 1976.

Jesse L. Lynn, Jr.
Lever Brothers Company

DETONATING AGENTS

. See EXPLOSIVES.